Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Arthur Kocher
le vendredi 23 juin 2017

Titre :

Biodiversité et maladies infectieuses: Impact des activités humaines sur le cycle de transmission des leishmanioses en Guyane

---

École doctorale et discipline ou spécialité**:**
ED SEVAB : Écologie, biodiversité et évolution

Unité de recherche :
Laboratoire EDB (UMR 5174 CNRS/UPS/ENFA)

Directeur/trice(s) de Thèse :

Jérôme Murienne
Anne-Laure Bañuls

Jury:
Rapporteurs:
Serge Morand et Brice Rotureau
Autres membres:
Jérôme Chave, Jean-François Guégan et Renaud Piarroux

REMERCIEMENTS

Ce travail de thèse a été rendu possible par l'implication, le conseil et le soutien d'un grand nombre de personnes et d'institutions.

Je tiens à remercier chaleureusement mes directeurs de thèse, Jérôme Murienne et Anne-Laure Bañuls. Jérôme, merci pour ta grande disponibilité et tes enseignements précieux. Merci également pour la confiance que tu m'as accordé depuis le début et pour la relation complice et constructive que nous avons eu. J'ai énormément appris de ces années de thèse. Anne-Laure, merci pour tes conseils avisés et pour avoir été là quand il le fallait même à distance.

Ce projet a été financé par le Labex CEBA, qui a également constitué un environnement scientifique stimulant et riche en opportunités de collaborations. D'importants soutiens apportés par le fond Inkermann (sous l'égide de la fondation de France), l'Observatoire Homme-Milieu Oyapock (Labex DRIIHM), le projet FEDER « RESERVOIR » et le laboratoire EDB ont permis l'extension du projet initial.

Je tiens particulièrement à remercier Jean-Charles Gantier, dont l'expertise taxonomique et l'implication sans faille ont rendu ce projet réalisable. Merci pour nous avoir accompagné dans nos premiers pas de « chasseurs » de phlébotomes et pour nous avoir communiqué tes connaissances et ta passion avec la bonne humeur et l'énergie qui te caractérisent.

Merci aux autres membres de l'ANR IAEL et du projet CEBA BIOHOPSYS, qui ont été à l'origine de ce projet, pour leur précieuse collaboration : François Catzeflis, Romain Girod, Pascal Gaborit et Ghislaine Prevost.

Merci à Benoit de Thoisy pour son aide indispensable dans la préparation et la réalisation de nos missions en Guyane et pour avoir permis la continuité du projet au-delà de cette thèse.

Merci également à Lucie Zinger, Cécile Cassan, Marine Ginouves, Pieternel Verschuren, Pierre-Michel Forget et Agathe Chavy pour leur participation à ce travail.

Merci à tous les collègues et amis du laboratoire EDB qui ont permis la réalisation de cette thèse dans des conditions privilégiées à tous les niveaux. Je mesure la chance que j'ai eu de me rendre au travail avec joie pendant plus de trois ans.

Enfin, merci à ma famille et à mes amis pour m'avoir apporté leur soutien dans les moments de doutes et de fatigue, et pour tout le reste évidemment !

PRÉAMBULE ET PRÉSENTATION DES OBJECTIFS

Dans le contexte actuel de changements environnementaux à grande échelle, des inquiétudes sont émises quant à la perte de services écosystémiques importants. L'intensification des phénomènes d'émergence de maladies d'origine animale dans une période qui connait un taux d'extinction d'espèce anormalement élevé soulève notamment des questions sur les conséquences sanitaires du déclin de la biodiversité. Les régions abritant une grande diversité d'organismes abritent également une grande diversité de pathogènes et sont donc classiquement considérées comme présentant un risque infectieux important. Pourtant, d'autres perspectives plus récentes envisagent la biodiversité comme un facteur de régulation des maladies.

Selon l'hypothèse de l'« effet de dilution », les écosystèmes les plus riches en espèces sont également les moins propices à la circulation des agents infectieux du fait de la présence d'hôtes non compétents constituant des impasses épidémiologiques. Si des travaux mathématiques et expérimentaux ont permis d'affirmer la validité théorique de ce mécanisme, la question de son importance effective en conditions naturelles reste controversée. Les études existantes se sont focalisées sur un nombre restreint de modèles emblématiques, et se basent souvent sur des preuves indirectes qui permettent difficilement la validation des processus sous-jacents. Ce débat scientifique souffre donc encore d'un manque d'éléments empiriques sur lesquels s'appuyer. Par ailleurs, la démonstration de la généralité du phénomène nécessite l'adoption d'approches multi-pathogènes qui sont encore trop rares. Il faut dire que la réalisation de telles études éco-épidémiologiques nécessite la collecte de données particulièrement difficiles à obtenir sur la prévalence des agents infectieux et la diversité des hôtes.

Les leishmanioses sont un ensemble de maladies causées par des parasites protozoaires du genre *Leishmania*, et transmises à l'homme ainsi qu'à d'autres mammifères par la piqure de petits diptères hématophages, les phlébotomes (Psychodidae : Phlebotominae). Elles se manifestent par différentes formes cliniques, provoquant des symptômes cutanés ou viscéraux. Les leishmanioses sont distribuées très largement dans les zones intertropicales et tempérées, et responsables de plus d'un million de cas chaque année. Sur le continent américain, la plupart des leishmanioses sont caractérisées par un cycle de transmission sylvatique impliquant différentes espèces de phlébotomes vecteurs et d'hôtes mammifères

sauvages. Bien qu'elles soient régulièrement qualifiées de maladies ré-émergentes, du fait de modifications environnementales telles que la déforestation, l'effet des perturbations d'origine humaine sur leur cycle de transmission n'a jamais été étudié sous l'angle de l'effet de dilution.

L'objectif principal de ce travail de thèse est de tester l'hypothèse de l'effet de dilution sur le cycle de transmission des leishmanioses cutanées zoonotiques en Guyane française. Les leishmanioses constituent un modèle d'étude multi-pathogènes, multi-vecteurs et multi-hôtes d'autant plus judicieux qu'elles représentent un enjeu sanitaire important. La Guyane française, couverte en grande majorité par un bloc continu de foret primaire, bénéficie encore aujourd'hui d'un statut de conservation favorable. Toutefois, la forte pression démographique et la recrudescence d'activités minières illégales que connait la région représentent une menace pour la biodiversité dans certaines zones. Ce territoire constitue l'un des rares "laboratoires naturels" existant pour étudier l'impact des activités humaines sur des écosystèmes relativement préservés.

Ce travail de thèse est constitué du développement d'outils moléculaires pour la levée de verrous méthodologiques existants, et de leur application à une étude écologique préliminaire. Il est présenté en quatre chapitres contenant des articles scientifiques publiés ou en cours de préparation.

L'établissement d'un cadre phylogénétique fiable constitue un socle important pour la compréhension des phénomènes écologiques. **Dans le premier chapitre**, nous employons la méthode de *genome skimming* afin de générer des données génomiques pour les phlébotomes du Nouveau Monde et d'apporter une contribution à la connaissance de leur histoire évolutive.

Dans le cas d'une maladie vectorielle comme la leishmaniose, il est nécessaire de connaitre la sensibilité des vecteurs hématophages à la disponibilité d'hôtes vertébrés dans le milieu pour envisager les conséquences globales d'un déclin de biodiversité. **Dans le second chapitre**, nous validons l'application du *metabarcoding* pour l'identification groupée de phlébotomes, permettant un gain d'efficacité considérable par rapport aux approches classiques. Nous employons ensuite la méthode pour étudier l'effet de modifications de la faune de vertébrés sur les communautés de phlébotomes.

La caractérisation de la diversité d'hôtes est indispensable à l'étude de l'effet de dilution. Dans le cas des vertébrés, celle-ci nécessite le déploiement de méthodes observationnelles contraignantes. **Dans le troisième chapitre**, nous évaluons la fiabilité d'un court marqueur mitochondrial pour l'identification taxonomique de mammifères par metabarcoding, et constituons une base de référence moléculaire quasiment exhaustive pour

les espèces présentes en Guyane. Nous employons ce marqueur pour l'analyse de repas sanguins de phlébotomes et de moustiques, et évaluons l'utilité de cette approche pour décrire les communautés de vertébrés sur les sites d'échantillonnage.

Dans le quatrième chapitre, nous développons une méthode d'identification des *Leishmania* spp. basée sur le séquençage haut-débit de fragments de minicercles kinétoplastiques amplifiés par PCR. La technique permet de combiner une capacité de détection élevée et une bonne fiabilité taxonomique. Nous employons la méthode pour estimer la prévalence de différentes espèces de *Leishmania* dans des phlébotomes collectés le long d'un gradient d'anthropisation dans la région de Saint-Georges de l'Oyapock. Les résultats sont mis en relation avec la caractérisation des communautés de phlébotomes et l'analyse de leurs repas sanguins sur les mêmes sites. Cette étude constitue une première exploration descriptive de l'effet de dilution sur les leishmanioses zoonotiques avant l'application du protocole à un échantillonnage plus large.

Ces quatre chapitres sont précédés d'une **introduction générale**. Nous commençons par présenter le cadre scientifique et socio-environnemental dans lequel s'inscrit notre travail. En particulier, nous proposons une revue de la recherche existante sur l'effet de dilution. Nous tentons ensuite de synthétiser des connaissances générales nécessaires à l'appréhension du modèle d'étude avant d'aborder plus précisément les aspects éco-épidémiologiques des leishmanioses en relation avec la question de thèse.

Le manuscrit s'achève sur une **discussion générale** dans laquelle nous résumons et discutons les résultats obtenus dans leur ensemble.

Note à propos de l'échantillonnage : Ce projet de thèse, dans sa forme initiale, devait s'appuyer sur la collecte d'un grand nombre de phlébotomes réalisée dans le cadre de l'ANR IAEL en 2007 et 2008. Ces échantillons ont pu être exploité en partie dans la phase de développement méthodologique, mais la majorité des résultats de biologie moléculaire obtenus ont constitué un échec, possiblement du fait de leur conservation trop longue. Cette constatation nous a conduits à mener deux missions de collecte sur le terrain, de façon tardive, dans les régions de Saint-Georges et d'Iracoubo (Counami). Les échantillons récoltés lors des ces missions ont permis l'obtention de bien meilleurs résultats, qui sont présentés le long de cette thèse. Ceci rappelle l'importance de travailler sur du matériel biologique de bonne qualité, même lorsque des outils puissants sont employés, et explique le nombre relativement restreint de données disponibles pour notre étude écologique. Il est à noter que les échantillons de Counami n'ont pas été exploités pour les deux derniers volets de la thèse. Du fait de la faible abondance des phlébotomes dans cette localité, près de dix fois moins

d'individus y ont été collectés qu'à Saint-Georges. Si cet échantillonnage à permis de caractériser les communautés de phlébotomes sur les sites (Chapitre 2), le très faible nombre de spécimens gorgés de sang et la quasi absence de *Leishmania* détectées n'ont pas permis de tirer de conclusions sur le cycle de transmission des parasites dans cette localité.

SOMMAIRE

INTRODUCTION GÉNÉRALE

1. MALADIES ÉMERGENTES DANS UN CONTEXTE DE CHANGEMENTS GLOBAUX

A. LE PHÉNOMÈNE D'ÉMERGENCE DE MALADIES

L'établissement de la théorie microbienne (Koch 1876), et les progrès qui s'en suivirent dans les sciences médicales, avec en particulier, le développement de la vaccination puis des antibiotiques, ont révolutionné notre capacité à combattre les agents pathogènes. Vers le milieu du 20ème siècle, les prédictions les plus optimistes annonçaient déjà l'éradication totale des maladies infectieuses (Cockburn 1963; Fauci 2001). Ces dernières sont pourtant, aujourd'hui encore, responsables de plus de 20% de la mortalité mondiale (GBD 2013 Mortality and Causes of Death Collaborators 2015). Bien qu'il soit en grande partie la conséquence d'une difficulté d'accès aux soins dans les régions les plus pauvres du monde (OMS 2012), ce constat révèle également notre impuissance face aux capacités adaptatives des agents pathogènes et à l'imprévisibilité des phénomènes infectieux.

Le terme de maladies infectieuses émergentes (MIE) est apparu dans les années 1950, mais sa définition précise a beaucoup évolué depuis (Rosenthal *et al.* 2015). Il est aujourd'hui généralement employé, tel que défini par l'Organisation Mondiale de la Santé (OMS), pour désigner « une maladie infectieuse nouvellement découverte ou qui connait une augmentation d'incidence, une expansion géographique, ou une modification du spectre d'hôtes ou de vecteurs » (www.who.int/zoonoses/emerging_zoonoses, consulté le 28/07/16). Les maladies émergentes représentent un problème sanitaire de longue date, comme peuvent en témoigner les différents récits de « peste » rapportés depuis l'antiquité (Morens *et al.* 2004).

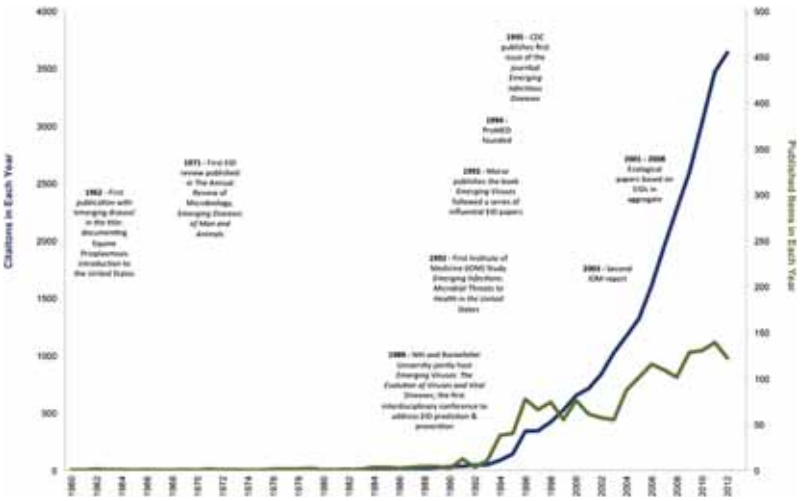


Figure 1: Evolution du nombre de publications (vert) et de citations (bleu) concernant les maladies infectieuses émergentes depuis 1960 (extrait de Rosenthal et al. 2015).

Bien que le poids des maladies infectieuses sur la santé mondiale ait tendance à diminuer avec le temps (GBD 2013 Mortality and Causes of Death Collaborators 2015), le siècle dernier a encore connu deux des épidémies les plus sévères de l'histoire. La pandémie de grippe de 1918 aurait fait environ 50 millions de morts, soit plus de victimes que la première guerre mondiale en quelques mois (Johnson & Mueller 2002). D'abord considéré comme un problème d'extension limitée lors de son apparition au début des années 80, le virus de l'immunodéficience humaine (VIH) et en particulier le syndrome d'immunodéficience acquise (sida) est progressivement devenu l'une des causes majeures de mortalité au début du 21ème siècle. Au total, le VIH a touché près de 70 millions de personnes dans le monde avec un taux de létalité proche de 50% (OMS, http://www.who.int/gho/hiv, consulté le 28/07/16).

Sans toujours être aussi dévastatrices, beaucoup d'autres maladies infectieuses ont émergé ou ré-émergé dans les dernières décennies, recevant un intérêt grandissant de la part de la communauté scientifique et des organisations sanitaires (Rosenthal *et al.* 2015, Figure 1). Parmi les exemples les plus notables, on peut citer l'apparition du virus Ebola, du syndrome respiratoire aigu sévère (SRAS) ou de la grippe aviaire H5N1, mais également la réémergence de la dengue, de la maladie de Lyme ou du Choléra (Figure 2). Par ailleurs, le phénomène n'est pas restreint aux maladies humaines: l'émergence de maladies infectieuses dans la faune sauvage et chez les plantes suscite des inquiétudes en matière de conservation et de sécurité alimentaire (Daszak 2000; Fisher *et al.* 2012).

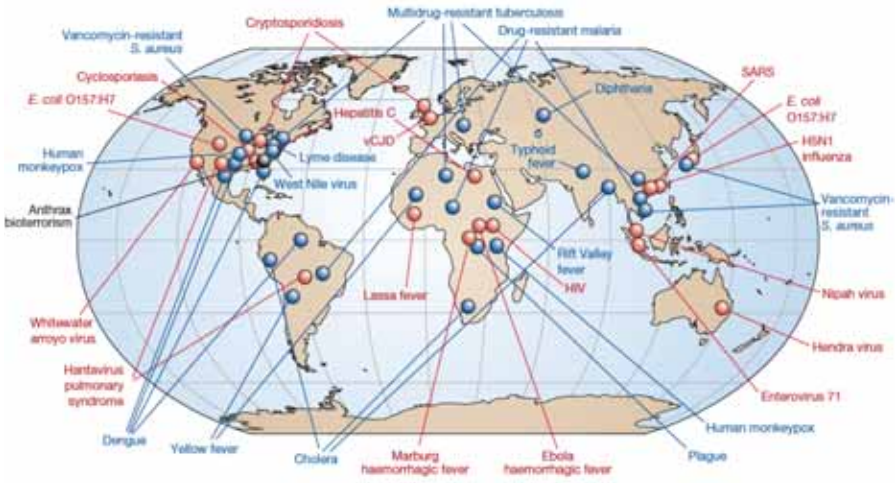


Figure 2: Exemples de maladies infectieuses émergentes (en rouge) et ré-émergentes (en bleu) (extrait de Morens *et al.* 2004).

En 2008, Jones *et al.* publient l'analyse spatio-temporelle d'un répertoire de 335 MIE. Bien que l'idée soit déjà courante dans la littérature scientifique (Morse 1995; Rosenthal *et al.* 2015), conduisant certains auteurs à considérer notre époque comme une nouvelle ère épidémiologique (Barrett *et al.* 1998; McMichael 2004), ce travail constitue la première démonstration quantitative de l'intensification du phénomène d'émergence de maladies infectieuses (Figure 3). Leurs résultats indiquent en effet que la fréquence d'apparition des MIE est en augmentation depuis les années 1940, même en prenant en compte un éventuel biais de déclaration. De plus, ils confirment les résultats d'études précédentes (Taylor *et al.* 2001; Woolhouse & Gowtage-Sequeria 2005) indiquant que la plupart de ces MIE sont causées par des pathogènes zoonotiques (60,3%), dont une majeure partie a pour origine la faune sauvage (71,8% des MIE zoonotiques).

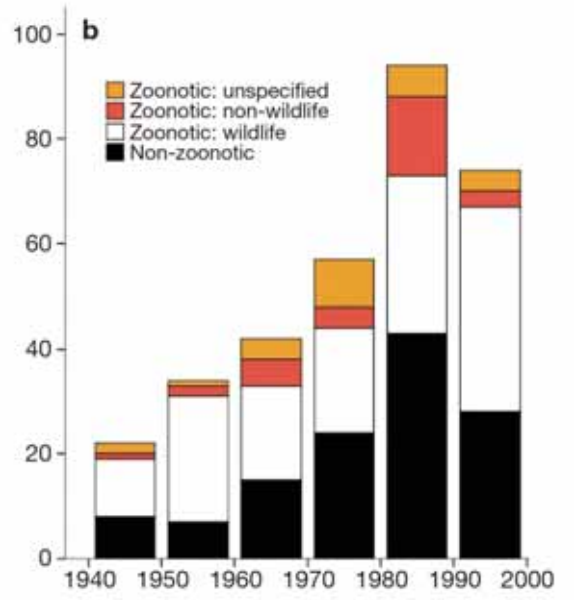


Figure 3: Nombre d'émergences de maladies infectieuses répertoriées par décennie depuis 1940, classées selon leur caractère zoonotique (extrait de Jones *et al.* 2008).

Un autre résultat important de leur étude est la détection d'une corrélation significative entre la densité de population humaine et la localisation géographique des émergences de maladies infectieuses. Ceci soutient une hypothèse déjà très répandue: les MIE sont, pour une grande part, la conséquence de changements démographiques et des activités humaines (Morse 1995; McMichael 2004; Woolhouse & Gowtage-Sequeria 2005).

B. LES FACTEURS ANTHROPOGENIQUES D'ÉMERGENCE DE MALADIES

Depuis la préhistoire, les évènements épidémiologiques majeurs sont liés à de grands changements socio-culturels ayant permis la mise en contact de pathogènes avec de nouvelles populations. L'abandon du mode de vie de chasseur-cueilleur et le développement de l'agriculture et de l'élevage il y a environ 10 000 ans se sont traduits par une densification des communautés et une proximité constante avec la faune nouvellement domestiquée. Cette situation aurait permis le passage de nombreux pathogènes d'origine animale à l'homme, chez qui certains circulent encore aujourd'hui (Weiss 2001). L'ouverture des contacts militaires et commerciaux entre l'Europe et l'Asie dans l'antiquité a conduit à des échanges de pathogènes à l'origine d'épidémies massives, telle que la peste de Justinien qui dévasta Constantinople (McNeill 1977). De façon similaire, lors de la conquête des Amériques, les colons européens ont traversé l'Atlantique accompagnés de leurs maladies infectieuses dont l'effet dévastateur sur les populations indigènes est aujourd'hui bien connu (Lovell 1992).

La croissance démographique exponentielle, le développement technologique et économique, les modifications radicales et la mondialisation des activités humaines que connait notre époque moderne constituent des bouleversements écologiques vraisemblablement impliqués dans la multitude de phénomènes infectieux émergents actuels (Jones *et al.* 2013).

Les changements d'utilisation des terres et l'invasion des milieux naturels sont souvent désignés parmi les principaux facteurs de risque, en particulier pour les MIE zoonotiques (Woolhouse & Gowtage-Sequeria 2005; Greger 2007). La population humaine est passée de 1 milliard au début du 20ème siècle à environ 7,5 milliards aujourd'hui, et comprendra probablement près de 10 milliards d'individus d'ici 2050 (Gonzalo *et al.* 2016). Cet accroissement exponentiel s'accompagne d'une lourde pression environnementale. La surface des forêts tropicales a été réduite de moitié au 20ème siècle (Myers *et al.* 1985), et la déforestation continue à un taux annuel de 2-3% (Patz *et al.* 2004). Plus d'un tiers de la production de biomasse terrestre est contrôlée par l'homme, et l'utilisation des terres à des fins de culture ou d'élevage continue à s'intensifier et à s'étendre, avec une surface totale qui devrait encore augmenter de près de 20% dans la première moitié du 21ème siècle (Tilman *et al.* 2001).

La déforestation peut constituer un changement environnemental favorable à la transmission de certaines maladies. Elle a par exemple coïncidé avec une recrudescence du

paludisme en Afrique, en Asie et en Amérique Latine, probablement par la création de sites de retenue d'eau stagnante favorables au développement des larves d'anophèles de manière plus abondante qu'en foret intacte (Patz *et al.* 2004). Par ailleurs, la déforestation et la construction de routes permettent la colonisation du milieu et la mise en contact de l'homme avec la faune sauvage, au travers d'activités telles que l'écotourisme, la chasse ou l'exploitation forestière (Wolfe *et al.* 2000). Ainsi, l'intensification de la chasse et de la consommation de viande de brousse en Afrique Centrale qui a eu lieu au cours du 20ème siècle est vraisemblablement liée à l'émergence de diverses zoonoses, dont la récente épidémie d'Ebola (Greger 2007).

Les pratiques de cultures modernes permettent parfois la prolifération d'hôtes ou de vecteurs de maladies déjà présentes. Des épidémies de fièvre hémorragique en Corée (virus Hantann) et en Argentine (virus Junin) ont résulté de l'augmentation de l'abondance des rongeurs porteurs de ces pathogènes suite à la mise en place de monocultures de riz ou de maïs (Morse 1995). L'expansion de l'encéphalite japonaise en Asie du Sud-Est a été associée à une intensification de la culture de riz irriguée qui favorise la pullulation des moustiques vecteurs du genre *Culex*, conjointement avec l'élevage de porc chez qui le virus circule, permettant ainsi son amplification et le passage à l'homme (Hurk *et al.* 2009; Jones *et al.* 2013). Le virus Hendra, en Australie, et le virus Nipah, en Malaisie, sont présents chez des chauves-souris frugivores du genre *Pteropus.* Leur récente émergence chez l'homme, provoquant des encéphalites mortelles, a été imputée à l'expansion de l'anthropisation dans les zones d'habitat naturel des chauve-souris, et à la présence, ici encore, d'une population animale domestique homogène et sensible au virus (le cheval pour le virus Hendra, et le porc pour le virus Nipah) ayant permis son amplification et son passage à l'homme (Plowright *et al.* 2014). L'émergence de la grippe aviaire en Asie serait également liée aux pratiques d'élevage. L'élevage de canards en plein air permet la mise en contact des volailles domestiques avec les oiseaux sauvages réservoirs du virus, les élevage intensifs en favorisent l'amplification, et l'existence de grands marchés vivants en permet la dissémination rapide (Jones *et al.* 2013).

La mondialisation des échanges commerciaux et l'intensification du trafic planétaire constituent également d'importants facteurs d'émergence de maladies en permettant la circulation des pathogènes sur de longues distances et leur mise en contact avec des populations naïves (Figure 4). Le VIH a pour origine un virus présent chez différentes espèces de primates non-humains qui est vraisemblablement passé chez l'homme suite à une morsure ou au contact avec du sang infecté lors de la chasse ou de la manipulation de viande de

brousse. On estime aujourd'hui que cet évènement est survenu à de multiple reprises depuis le début du siècle dernier, et que le virus circulait chez l'homme de manière localisée en Afrique bien avant le déclenchement de la pandémie dans les années 80 (Locatelli & Peeters 2012). C'est probablement l'intensification des mouvements de population des milieux ruraux reculés vers les grandes villes qui a permis la dissémination du virus jusqu'alors confiné dans des zones restreintes. L'épidémie de syndrome respiratoire aigu sévère (SRAS) survenue en 2003 est un exemple emblématique du rôle des transports longue distance et de la mondialisation dans la propagation rapide de pathogènes. D'abord apparue dans la province du Guangdong en Chine, la maladie s'est ensuite déclarée à Hong Kong, dans un hôtel de voyageurs, permettant sa dissémination rapide dans le monde entier. En quelques mois, le SARS a touché plus de 8000 personnes et causé 774 morts dans 26 pays et sur cinq continents (Peiris *et al.* 2003).

Figure 4: Représentation du trafic aérien actuel. Les lignes représentent les liens directs entre aéroports, et les couleurs indiquent la capacité de transport de passagers (croissante du bleu au rouge) (extrait de Kilpatrick & Randolph 2012).

Le cycle de transmission d'un grand nombre de pathogènes est dépendant des conditions climatiques (Harvell *et al.* 2002). Le réchauffement planétaire, induit essentiellement par l'accroissement des émissions de gaz à effet de serre lié aux activités humaines, peut donc avoir des conséquences épidémiologiques. Bien qu'il s'agisse d'un sujet controversé, et que le rôle des changements climatiques dans l'émergence de pathogènes soit aujourd'hui généralement considéré comme mineur en comparaison d'autres facteurs (Hay *et al.* 2002; Rohr *et al.* 2011; Kilpatrick & Randolph 2012), des modifications dans la prévalence et la sévérité de maladies infectieuses liées aux changements climatiques ont tout de même été identifiées (Harvell *et al.* 2002; Altizer *et al.* 2013).

D'une part, la hausse des températures peut avoir des effets sur le taux de multiplication des pathogènes et sur l'immunité des vecteurs ou des hôtes. L'augmentation des fluctuations de températures (l'une des composantes du changement climatique) a par exemple été associée à une diminution de la résistance des grenouilles au pathogène fongique *Batrachochytrium dendrobatidis*, participant ainsi à leur déclin en Amérique Latine.

D'autre part, les changements climatiques peuvent avoir des conséquences sur la distribution et la phénologie des espèces d'hôtes et de vecteurs de maladies. La potentielle expansion géographique du moustique tigre (*Aedes albopictus*, vecteur de la Dengue et du Chikungunya), permise par les hausses de température, suscite des craintes en Europe et aux Etats-Unis (Ruiz-Moreno *et al.* 2012; Rochlin *et al.* 2013; Campbell *et al.* 2015, Figure 5). Le déplacement vers le nord des populations de tiques du genre *Ixodes* a été imputé au réchauffement climatique (Lindgren *et al.* 2000; Ogden *et al.* 2006), et pourrait être à l'origine de l'émergence de la maladie de Lyme au Canada et de l'encéphalite à tique en Suède (Lindgren & Gustafson 2001; Ogden *et al.* 2009).

L'atténuation du froid hivernal dans certaines régions peut également diminuer les restrictions saisonnières préexistantes dans le cycle de transmission des pathogènes. Une étude conduite chez le papillon Monarque (*Danaus plexippus*), qui migre typiquement du nord vers le sud des Etats-Unis pour la période hivernale, montre que la hausse des températures et la culture de fruits exotiques ont permis la sédentarisation des populations dans certaines régions, entrainant une augmentation de la prévalence du parasite *Ophyrocystis elektroscirrha* (Altizer *et al.* 2011).

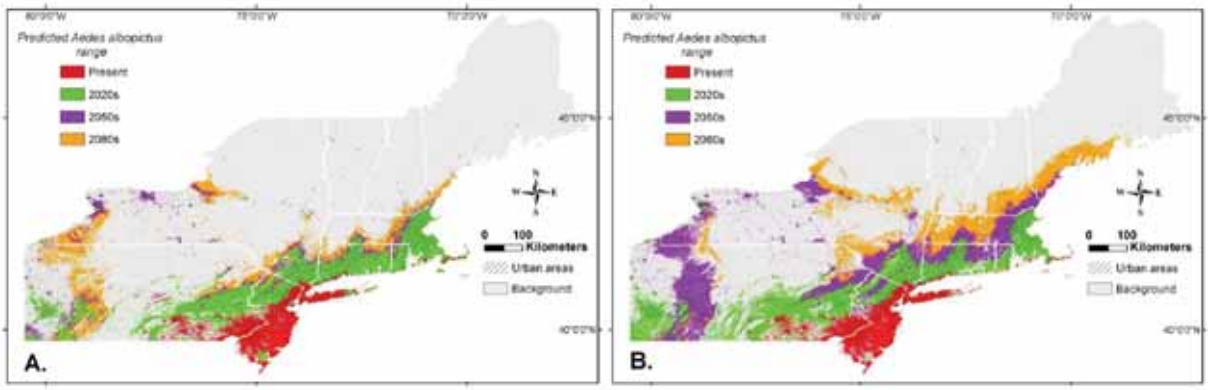


Figure 5: Prédictions de l'expansion géographique du moustique tigre (*Aedes albopictus*) dans le Nord-est des Etats-Unis selon deux scénario: **(A)** Augmentation modéré des émissions de CO_2, **(B)** augmentation élevée des émissions de CO_2 (extrait de Rochlin *et al.* 2013).

La démonstration du rôle majeur des changements environnementaux induits par l'homme dans l'émergence de ces maladies zoonotiques appelle à une meilleure prise en compte du contexte écologique pour l'étude des problématiques sanitaires. Le développement récent des mouvements EcoHealth et One Health témoigne de cette volonté au niveau international (Cook *et al.* 2004; Zinsstag 2012). Dans ce contexte, l'écologie des maladies, ou « éco-épidémiologie », est un champ disciplinaire grandissant (Kilpatrick & Altizer 2010).

2. LA BIODIVERSITÉ RÉGULATRICE DE MALADIES : HYPOTHÈSE DE L'EFFET DE DILUTION

L'une des composantes majeures des changements globaux actuels est l'érosion de la biodiversité. On estime que le taux d'extinction d'espèces de vertébrés est 100 fois plus élevé aujourd'hui qu'avant la propagation planétaire de l'homme, actuellement responsable de ce que certains qualifient de sixième extinction de masse (Ceballos *et al.* 2015). La dégradation concomitante des écosystèmes soulève des inquiétudes quand à la perte de fonctions essentielles pour le bien être des populations humaines (Cardinale *et al.* 2012). En particulier, la littérature concernant ces services écosystémiques mentionne régulièrement la biodiversité comme un facteur de régulation des maladies, au travers d'un mécanisme connu sous le nom d'« effet de dilution ».

A. DÉVELOPPEMENT DU CONCEPT ET PREMIERS ÉLÉMENTS EMPIRIQUES : LE CAS DE LA MALADIE DE LYME

Une vision classique des relations qui existent entre biodiversité et maladies infectieuses est que les zones abritant une grande diversité d'organismes abritent également une grande diversité de pathogènes, et représentent donc un risque sanitaire élevé. Il a été montré que la diversité de pathogènes humains est corrélée géographiquement à la richesse spécifique de vertébrés (Dunn *et al.* 2010, Figure 6a), suivant ainsi le gradient latitudinal de biodiversité classiquement décrit (Guernier *et al.* 2004, Figure 6b). Par ailleurs, la diversité d'espèces de mammifères sauvages a été identifiée comme un prédicteur significatif de l'émergence de maladies zoonotiques (Jones *et al.* 2008).

Une perspective différente attribue à la biodiversité un rôle de régulation des maladies. Bien qu'elle ait suscité un fort intérêt récemment, cette idée n'est pas nouvelle. L'avantage de la polyculture pour limiter l'impact des pestes agricoles est en effet connu depuis longtemps (Altieri & Nicholls 2004). L'idée que la biodiversité pourrait représenter un facteur de protection contre les maladies infectieuses zoonotiques a en revanche émergée plus récemment, *via* des études pionnières sur la maladie de Lyme.

La maladie de Lyme est une zoonose causée par une bactérie spirochète (*Borrelia burgdorferi*) et transmise par des tiques du genre *Ixodes* dans tout l'hémisphère nord où elle est aujourd'hui en expansion (Randolph 2001; Steere *et al.* 2004). Au début des années 1990,

des chercheurs mettent en évidence que certaines espèces de vertébrés (oiseaux et ongulés), sur lesquels se nourrissent les tiques, sont beaucoup moins compétentes que d'autres à acquérir et transmettre la bactérie (Matuschka & Spielman 1992; Matuschka *et al.* 1993). Chaque morsure de tique sur l'une de ces espèces constitue un « gâchis » épidémiologique qui entrave le cycle de transmission de la maladie. La présence de ces hôtes peu compétents dans les écosystèmes peut donc contribuer à diminuer le risque d'infection et représenter un bénéfice pour la santé publique. Ce phénomène n'est pas sans rappeler le bien antérieur principe de zooprophylaxie, selon lequel l'attraction des vecteurs du paludisme par des animaux domestiques non compétents (typiquement le bétail) constitue un moyen de lutte contre la maladie (Macdonald 1956).

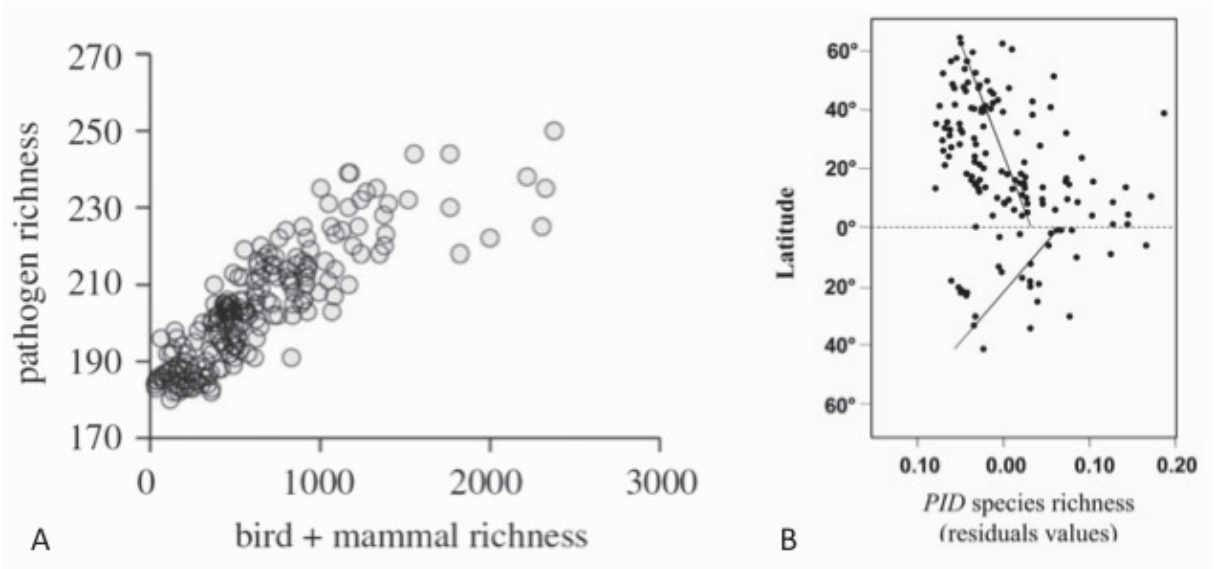


Figure 6: (A) Richesse spécifique de pathogènes en fonction de la richesse spécifique de vertébrés à l'échelle mondiale (extrait de Dunn *et al.* 2010). **(B)** Richesse spécifique des agents infectieux et parasitaires en fonction de la latitude (extrait de Guernier *et al.* 2004).

Ce mécanisme fut ensuite exploré par Van Buskirk & Ostfeld (1995) avec un modèle épidémiologique indiquant que la présence d'un hôte non compétent peu contribuer à la diminution du risque de transmission dans la mesure où il ne contribue pas trop à l'accroissement de la population de tiques. Le terme « effet de dilution » fut employé pour la première fois dans une étude modélisant l'efficacité de la transmission du virus Louping ill en fonction de la densité de deux hôtes dont l'un est compétent et l'autre ne l'est pas (le Lagopède d'Ecosse et le Lièvre variable, Norman *et al.* 1999). Le mécanisme a par la suite été étudié plus précisément au travers de diverses études centrées sur la maladie de Lyme dans le Nord-est des Etats-Unis. Schmidt & Ostfeld (2001) puis LoGiudice *et al.* (2003) développent des modèles basés sur l'estimation empirique de paramètres de transmission, et observent des taux de prévalence de la bactérie dans les nymphes de tiques collectées sur le terrain très

inférieurs aux taux attendus si celles-ci se nourrissaient uniquement sur les hôtes les plus compétents: la souris à pattes blanches (*Peromyscus leucopus*) et le tamia rayé (*Tamias striatus*, Figure 7a). Ils apportent ainsi des éléments quantitatifs pour démontrer l'importance de prendre en compte les hôtes alternatifs non-compétents dans l'étude des maladies vectorielles.

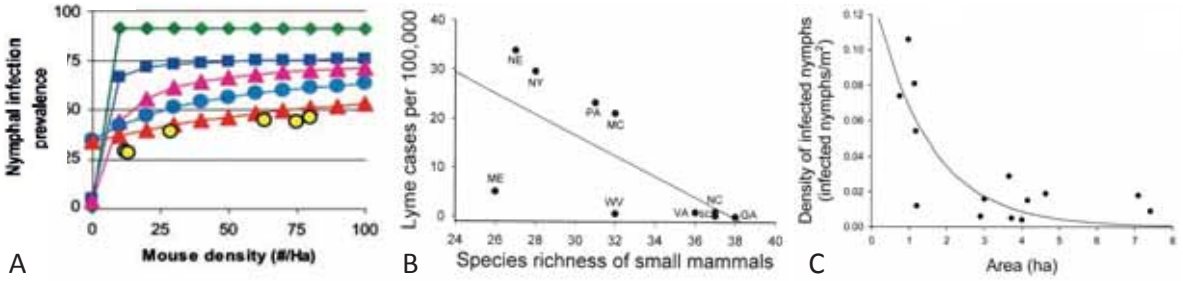


Figure 7: (A) Modèle épidémiologique de la maladie de Lyme basé sur l'estimation empirique de paramètres de transmission. Le taux d'infection dans les nymphes de tiques prévu par le modèle est présenté en fonction de la densité du réservoir principal (la souris à pattes blanches), et en retirant progressivement des hôtes non-compétents du système (du triangle rouge avec l'écosystème complet aux losanges verts ou seule la souris à pattes blanche est présente). Les ronds jaunes représentent des mesures empiriques collectées sur sept ans (extrait de LoGiudice *et al.* 2003). **(B)** Nombre de cas de maladie de Lyme dans différents états de l'est des Etats-Unis en fonction de la richesse spécifique de petits mammifères (extrait de Ostfeld & Keesing 2000a). **(C)** Abondance des nymphes de tiques infectées par la maladie de Lyme en fonction de la taille du fragment forestier considéré (extrait de Allan *et al.* 2003).

Des études empiriques corrélatives furent menées en parallèle afin d'évaluer l'impact de l'effet de dilution le long de gradients de biodiversité, partant de la constatation que l'hôte le plus compétent pour la maladie (*Peromyscus leucopus*) est typiquement surreprésenté dans les forêts perturbées abritant une faible diversité de vertébrés. Ostfeld & Keesing (2000) montrent l'existence d'une corrélation négative entre le taux de prévalence chez l'homme et la diversité de petits mammifères à l'échelle des régions dans l'est des Etats-Unis (Figure 7b). A une échelle plus locale, Allan *et al.* (2003) mettent en évidence une corrélation négative entre la densité de nymphes infectées et la taille des fragments forestiers, dont les plus petits sont supposés abriter une plus faible biodiversité, et une plus grande densité de souris à pattes blanches (Nupp & Swihart 1998, Figure 7c). Cette étude constitue un premier élément empirique associant l'altération de la biodiversité par l'homme (causée ici par la fragmentation forestière) à une augmentation du risque de transmission d'une maladie zoonotique.

B. GÉNÉRALISATION

Ces travaux portant sur la maladie de Lyme entraient en résonnance avec l'essor de la recherche sur les services écosystémiques (Loreau *et al.* 2001) et suscitèrent de nouveaux

questionnements concernant l'impact de l'altération de la biodiversité sur le risque de transmission des maladies infectieuses. Plusieurs études ont testé l'hypothèse de l'effet de dilution pour d'autres maladies vectorielles (Gilbert *et al.* 2001; Telfer *et al.* 2005; Vaz *et al.* 2007). En particulier, différents travaux mirent en évidence une corrélation négative entre la diversité d'espèces d'oiseaux et la prévalence du virus du Nil occidental sur de larges échelles spatiales aux Etats-Unis (Ezenwa *et al.* 2006; Swaddle & Calos 2008; Allan *et al.* 2009).

Des études furent aussi conduites sur des pathogènes à transmission directe, chez les plantes en particulier. Roscher *et al.* (2007), par exemple, constituent différents assemblages végétaux et constatent que les assemblages les plus riches en espèces sont les moins affectés par les pathogènes fongiques du genre *Puccinia*. Chez les animaux, plusieurs travaux s'intéressèrent au cas des hantavirus. Ces virus circulent chez des hôtes réservoirs rongeurs et peuvent infecter l'homme suite au contact avec des fèces, de la salive ou de l'urine contaminée mais aussi par l'inhalation d'aérosols. Ruedas *et al.* (2004) montrent que les sites ayant connu des épidémies d'hantavirus au Panama sont caractérisés par une faible diversité de petits mammifères, et une surreprésentation des espèces de rongeurs réservoirs, probablement due à la culture de la canne à sucre et du maïs. L'augmentation de l'abondance des rongeurs réservoir et de la prévalence d'hantavirus fut également constatée au Mexique suite à l'extirpation expérimentale d'espèces de petits mammifères non compétents (Suzán *et al.* 2009).

L'abondance d'un hôte réservoir peut être limitée par la présence d'autres espèces du fait de la compétition interspécifique, suffisant à expliquer l'existence d'une relation négative entre richesse spécifique et prévalence d'un pathogène à transmission densité dépendante (Begon 2008). Ce mécanisme peut être considéré comme un deuxième type d'effet de dilution, distinct de celui qui fut originalement décrit pour les maladies vectorielles (Rudolf & Antonovics 2005). Plus récemment, d'autres études portant sur les hantavirus indiquèrent un effet de la diversité *per se* sur la circulation du virus, n'étant pas limité à la diminution de la densité de l'hôte réservoir (Dizney & Ruedas 2009; Clay et al. 2009a). Ces travaux semblent révéler un mécanisme lié à des modifications comportementales de l'hôte en présence d'autres espèces, une hypothèse soutenue par l'observation d'une réduction des contacts intra spécifiques chez la souris sylvestre sur des sites caractérisés par une plus grande diversité et une prévalence du VSN inférieure (Clay *et al.* 2009b; Dizney & Dearing 2016). Les auteurs estiment que leurs résultats peuvent s'expliquer par la sélection de profils comportementaux moins explorateurs due à des interactions interspécifiques telles que la prédation.

L'impact des modifications de biodiversité fut également étudié pour le cas de maladies parasitaires à transmission indirecte incluant un stade de développement libre. Johnson *et al.* (2008) montrent expérimentalement que des larves de crapaud américain exposées au trématode *Ribeiroia ondatrae* sont moins fréquemment affectées lorsqu'elles sont en présence de larves de rainettes versicolores peu sensibles au parasite. Une étude similaire fut menée sur le trématode *Shistosoma mansoni*, agent de la bilharziose intestinale chez l'homme (Johnson *et al.* 2009). Les résultats ont montré un taux d'infestation réduit chez l'hôte intermédiaire *Biomphalaria glabrata* en présence d'autres espèces de gastéropodes. Ces expériences, dans lesquelles la densité d'hôte est contrôlée, mettent en évidence un type d'effet de dilution encore distinct que les auteurs associe à l' « effet leurre » (*decoy effect*) décrit bien antérieurement (Chernin & Perlstein 1971; Combes & Moné 1987). L'effet leurre désigne l'interférence d'espèces non-hôtes avec l'infestation des espèces hôtes par la forme libre d'un parasite. Il s'explique en particulier par la provocation de déplacements supplémentaires menant à l'épuisement du parasite ou par l'occurrence d'infection des hôtes incompatibles ne permettant pas la poursuite du cycle de développement.

Face à cette multitude d'études, portant sur divers systèmes épidémiologiques, Keesing *et al.* (2006), puis Keesing *et al.* (2010) redéfinissent le concept d'effet de dilution et tentent de détailler le cadre de travail associé. Faisant à l'origine spécifiquement référence au mécanisme de déroutement de vecteurs de pathogènes sur des hôtes alternatifs non-compétents, les auteurs proposent d'élargir la définition de l'effet de dilution pour désigner le phénomène de régulation des maladies infectieuses par la biodiversité d'une manière plus générale. Ils dressent une liste de mécanismes, présentés comme différents « modes » de l'effet de dilution, par le biais desquels la présence d'hôtes non-compétents peut impacter les paramètres de transmission (Tableau 1). Keesing *et al.* (2010) précisent qu'en théorie, une diminution de la biodiversité peut favoriser ou réduire la transmission, et définissent des conditions, précisées plus tard par Ostfeld & Keesing (2012), pour lesquelles un effet de dilution est attendu: (i) les hôtes diffèrent dans leur compétence à transmettre le pathogène ou à nourrir les vecteurs, (ii) les hôtes les plus compétents ont tendance à être présents dans les communautés les plus pauvres en espèces (iii), la présence d'hôtes non-compétents régule l'abondance des hôtes compétents ou des vecteurs, ou réduit les taux de contact hôte-pathogène ou hôte-vecteur (Encadré 1). Ils suggèrent que ces conditions sont effectivement rencontrées fréquemment, expliquant le grand nombre d'exemples de pathogènes dont la transmission est favorisée par une diminution de la biodiversité.

Encadré 1: Conditions pour lesquelles un effet de dilution est attendu (d'après Ostfeld & Keesing 2012).

1. Les hôtes diffèrent dans leur compétence à transmettre le pathogène ou à nourrir les vecteurs.

2. Les hôtes les plus compétents ont tendance à être présents dans les communautés les plus pauvres en espèces, tandis que les hôtes les moins compétents ont tendance à être présents dans les communautés les plus diverses.

3. La présence d'hôtes non-compétents régule l'abondance des hôtes compétents ou des

Cette généralisation suscite débat. Randolph & Dobson (2012) critiquent lourdement l'idée selon laquelle la préservation de la biodiversité permet de réguler la transmission des maladies infectieuses. Ils jugent que l'hypothèse de l'effet de dilution résulte d'une volonté de justifier des objectifs de conservations par des arguments utilitaires palpables (ainsi qu'un moyen efficace d'attirer des fonds de recherche). Ils dénoncent des travaux de modélisation mathématique basés sur des hypothèses fortes et hautement spéculatives, des études empiriques peu concluantes, et fournissent une série de contre-exemples. Wood & Lafferty (2013) rappellent que la réémergence de la maladie de Lyme dans le nord-est des Etats-Unis est vraisemblablement liée à la reforestation et au retour de la faune sauvage suite au déplacement de l'agriculture dans d'autres régions. Ils ajoutent que l'élimination de la végétation et des hôtes réservoirs est considérée par les épidémiologistes comme le moyen de lutte le plus efficace contre cette maladie. Ils appellent à plus de nuance concernant les potentielles vertus protectrices de la biodiversité, qu'ils considèrent très dépendante de l'échelle spatiale et du niveau d'anthropisation considérés. Salkeld *et al.* (2013) conduisent une meta-analyse de 16 études dont les résultats suggèrent une relation négative très faible et hétérogène entre biodiversité et risque infectieux, ne permettant pas d'envisager l'effet de dilution comme un phénomène général et significatif pour la santé publique. De façon similaire, Wood *et al.* (2014) émettent des prédictions qualitatives de l'effet de la biodiversité pour une liste de 69 maladies parasitaires ayant un impact sanitaire significatif, en fonction des caractéristiques écologiques des hôtes et des vecteurs. Ils concluent que dans la plupart des cas, la biodiversité ne devrait pas affecter le risque infectieux, et que les cas de corrélation négative entre biodiversité et risque infectieux sont moins fréquents que l'inverse. Cependant, une meta-analyse plus récente incluant 202 études conclue à un effet négatif significatif de la biodiversité sur l'abondance des pathogènes, quelque soit le type de pathogène ou le mode de transmission considérés (Civitello *et al.* 2015). La méthode employée ne permet toutefois pas de détecter un éventuel biais de publication et inclue une grande proportion d'études

expérimentales dans lesquelles la structure de la communauté est manipulée de manière artificielle.

Cette polémique concernant la généralité de l'effet de dilution constitue une mise en garde contre la proclamation précoce et démesurée des vertus « anti-infectieuses » de la biodiversité. A l'heure actuelle, la conservation des milieux naturels ne peut pas être considérée comme un instrument prophylactique efficace, et l'adoption d'une attitude scientifique nuancée est importante à l'égard des politiques de santé publique (Salkeld *et al.* 2015). Il reste néanmoins fondamental d'étudier les conséquences diverses des activités humaines sur le fonctionnement des écosystèmes.

Tableau 1: Liste de mécanismes potentiels par lesquels la présence d'hôtes non-compétents pour un pathogène peut conduire à la diminution (effet de dilution) ou à l'augmentation (effet d'amplification) de sa transmission (d'après Keesing *et al.* 2006).

Paramètre de transmission	**Dilution**	**Amplification**
Abondance de l'hôte	La présence d'autres espèces limite l'abondance de l'hôte via la compétition pour les ressources ou la prédation.	
Abondance des vecteurs	La présence de certaines espèces conduit à une augmentation de la mortalité des vecteurs par des comportements de défense ou des réactions immunitaires.	La présence d'espèces supplémentaires fournit plus d'opportunités de repas sanguins aux vecteurs conduisant à une augmentation de leur abondance.
Taux de contact entre les hôtes	Le taux de contact intra-spécifique est diminué par la présence d'autres espèces: la présence de prédateurs limite les comportements exploratoires, le nombre d'interactions par individu est fixe, les interactions interspécifiques entrainent une diminution des interactions intra-spécifiques.	Le taux de contact intra-spécifique est augmenté par la présence d'autres espèces, si elle entraine par exemple une agrégation plus dense des hôtes.
Taux de contact entre le vecteur et l'hôte	Le vecteur est généraliste et la présence d'espèces non-compétentes diminue la fréquence des repas sanguins pris sur l'hôte compétent. La présence d'autres espèces induit des changements comportementaux qui diminuent les taux de contacts entre l'hôte et le vecteur.	La présence d'autres espèces induit des changements comportementaux qui augmentent les taux de contacts entre l'hôte et le vecteur.
Susceptibilité de l'hôte	La présence d'autres espèces proies ou mutualistes favorise les capacités immunitaires de l'hôte.	La présence d'espèces compétitrices ou prédatrices augmente le stress et affaiblit la réponse immunitaire de l'hôte.

C. APPROFONDISSEMENT THÉORIQUE : PERSPECTIVES ÉCOLOGIQUES ET ÉVOLUTIVES

L'intérêt récent pour les questions reliant biodiversité et maladies infectieuses s'accompagne encore de faiblesses théoriques et d'un manque de données de terrain sur lesquelles s'appuyer. Parmi les 202 études répertoriées par Civitello *et al.* (2015), seules 34 sont basées sur des observations de terrain (et non pas sur des manipulations expérimentales). La difficulté à caractériser directement et simultanément la diversité de la faune sauvage et la prévalence d'agents infectieux explique probablement le nombre limité d'études réalisées en conditions naturelles, bien que celles-ci soient essentielles à la validation d'hypothèses écologiques. Par ailleurs, du fait de son développement récent et de sa nature transdisciplinaire, ce champ de recherche souffre encore d'un manque d'homogénéité lexical et conceptuel.

La biodiversité, un terme désignant la diversité de la vie sous toutes ses formes, revêt plusieurs facettes et peut être quantifiée par différents indices. Des indices prenant en compte l'abondance relative des espèces (*e.g.* l'indice de Shannon) paraissent plus adaptés que la richesse en espèce brute pour étudier l'effet des modifications dans la structure des communautés sur la circulation des pathogènes. Par ailleurs, l'emploi de la diversité phylogénétique ou fonctionnelle plutôt que de la diversité taxonomique peut constituer une approche judicieuse, notamment pour l'étude des pathogènes multi-hôtes. Par exemple, Parker *et al.* (2015) montrent expérimentalement que la sévérité et la prévalence des infections chez une espèce végétale sont mieux prédites lorsque que l'abondance des espèces phylogénétiquement proches est considérée.

De façon similaire, les mesures de risque infectieux employées ne sont pas toujours les mêmes selon les études, variant entre prévalence absolue chez les hôtes ou les vecteurs, taux de prévalence, diversité parasitaire ou nombre de cas humains. Ces confusions portent sur des aspects pratiques de la recherche, mais révèlent des problèmes de cohérence plus fondamentaux dans la définition des problématiques.

Les raccourcis sémantiques employés dans la littérature peuvent laisser croire à l'existence d'une théorie générale de « l'impact de la biodiversité sur l'émergence et la transmission des maladies infectieuses » (Keesing *et al.* 2010). Il faut pourtant distinguer les questions d'ordre biogéographique, visant à expliquer les patrons de diversité parasitaire à l'échelle planétaire, de celles portant sur les conséquences de perturbations locales des écosystèmes sur le cycle de transmission des pathogènes (Figure 8a).

De la même manière, perturbations anthropogéniques et déclin de biodiversité sont souvent considérés comme de façon indissociable, l'effet de dilution permettant de faire le lien entre ces perturbations et une modification du risque infectieux. Or, les perturbations d'origine humaine peuvent impacter la transmission des agents pathogènes par d'autres mécanismes, et elles s'accompagnent parfois d'une augmentation de la richesse en espèces locale (Sax & Gaines 2003; Young *et al.* 2013). Il est donc important de distinguer la problématique générale des mécanismes sous-jacents (effet de dilution, entre autre).

Un raisonnement par l'absurde permet de se rendre compte de l'impossibilité d'une relation négative constante entre biodiversité et risque infectieux. Dans un environnement hypothétique totalement anthropisé, ou la faune sauvage est inexistante (« l'écosystème du parking »; Levi *et al.* 2016), la prévalence des pathogènes zoonotiques est forcément nulle. Si un effet de dilution se manifeste, c'est donc au niveau d'une fenêtre réduite le long d'un gradient d'anthropisation allant de l'urbanisation totale à un milieu naturel intact (Figure 8b).

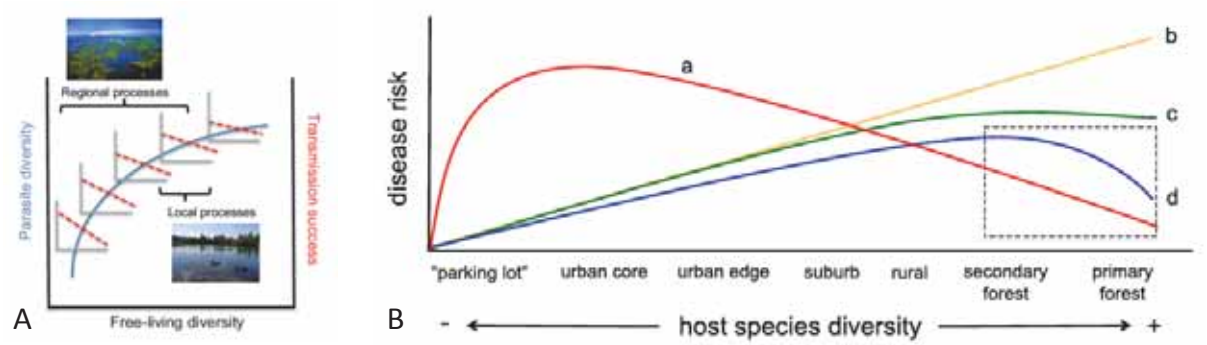


Figure 8: (A) Schématisation d'effet potentiellement opposé de la diversité d'hôtes sur la diversité parasitaire (en bleu) et le succès de transmission (en rouge) selon l'échelle spatiale considérée (extrait de Johnson *et al.* 2015). **(B)** Modèles théoriques de la relation existante entre le niveau d'anthropisation du milieu/la diversité d'hôtes et le risque infectieux. La prévalence des maladies zoonotiques est forcément nulle lorsque la diversité d'hôtes est nulle (l'« écosystème du parking »). Les relations possibles incluent a) un effet de dilution, b) un effet d'amplification, c) un effet d'amplification qui sature lorsque la biodiversité est élevée, d) un effet d'amplification qui s'inverse lorsque la biodiversité est élevée. La zone en pointillés indique comment la sélection d'une fourchette réduite dans le gradient et pour un scénario spécifique permet de conclure à l'existence d'un effet de dilution (extrait de Wood *et al.* 2016).

Différents mécanismes pouvant relier biodiversité et cycle de transmission des pathogènes ont été proposés (Tableau 1, Keesing *et al.* 2006). Toutefois, ceux-ci sont indépendants les uns des autres et peuvent avoir des effets opposés. Bien que des efforts aient été faits pour synthétiser les conditions dans lesquelles une augmentation de la biodiversité conduit à une diminution de la prévalence d'un pathogène particulier (Encadré 1), la généralisation de l'effet de dilution est supposée sur la base de l'accumulation (contestée) d'observations empiriques, sans en caractériser les fondements théoriques. Un progrès en ce sens serait de s'appuyer de manière plus importante sur les champs plus anciens de l'écologie

des communautés et de l'écologie évolutive pour, en particulier, (i) prédire les effet des perturbations anthropiques sur la structure des communautés via des règles d'assemblage non-aléatoires en regard de la transmission des agents pathogènes, (ii) adopter une perspective plus globale, multi-pathogène, considérant les organismes parasitaires comme des composantes à part entière des écosystèmes.

Une hypothèse fréquemment évoquée est l'existence d'une corrélation entre résilience des hôtes (et donc propension à dominer dans les habitats perturbés) et compétence pour les pathogènes. Les hôtes les plus résilients ont tendance à être caractérisés par de larges aires de répartitions géographiques, des taux de reproduction, des métabolismes et des abondances élevés, des tailles corporelles et des durées de vie réduites (Cardillo *et al.* 2008). Ces traits sont caractéristiques d'une « stratégie d'histoire de vie rapide », qui semble associée à une plus grande compétence pour les pathogènes: Huang *et al.* (2015) montrent que la compétence des hôtes pour la maladie Lyme et de l'encéphalite équine de l'Est diminue avec la taille et le temps de gestation/incubation, tandis qu'elle est positivement corrélée à la taille de portée/ponte. Chez les plantes, Cronin *et al.* (2010) montrent que les espèces ayant un fort métabolisme sont les plus compétentes pour le virus du nanisme jaune. Parmi 13 espèces d'amphibiens étudiées par Johnson *et al.* (2012), les plus sensibles au trématode *R. ondratae* se révèlent être celles qui présentent un développement rapide, une petite taille à la métamorphose et une durée de vie courte. Plus récemment Han *et al.* (2015), montrent que parmi 2277 espèces de rongeurs répertoriées, la probabilité de jouer le rôle de réservoir pour un pathogène zoonotique est largement prédite par une stratégie d'histoire de vie rapide et une distribution géographique large.

Ce lien pourrait s'expliquer par l'existence d'un compromis dans l'allocation des ressources, les espèces ayant un développement rapide et un taux de reproduction important investissant moins d'énergie dans la production de capacités immunitaires (Martin *et al.* 2006; Previtali *et al.* 2012). Par ailleurs, il est possible que l'adaptation aux hôtes les plus abondants, et donc, les plus fréquemment rencontrés, constitue un avantage sélectif pour les parasites (Johnson *et al.* 2015). Ces hypothèses prédisent que les espèces présentes dans les habitats les plus perturbés, abritant les niveaux de diversité les plus faibles, sont également les plus compétentes pour les pathogènes, une tendance qui a été récemment observée directement dans des communautés d'amphibiens exposées au trématode *R. ondratae* (Johnson *et al.* 2013, Figure 9). La littérature n'est toutefois pas unanime au sujet des liens existants entre traits d'histoire de vie, résilience et capacités immunitaires (Isaac &

Cowlishaw 2004; Cooper *et al.* 2012; Young *et al.* 2013). D'avantage d'études sont nécessaires pour soutenir ces hypothèses.

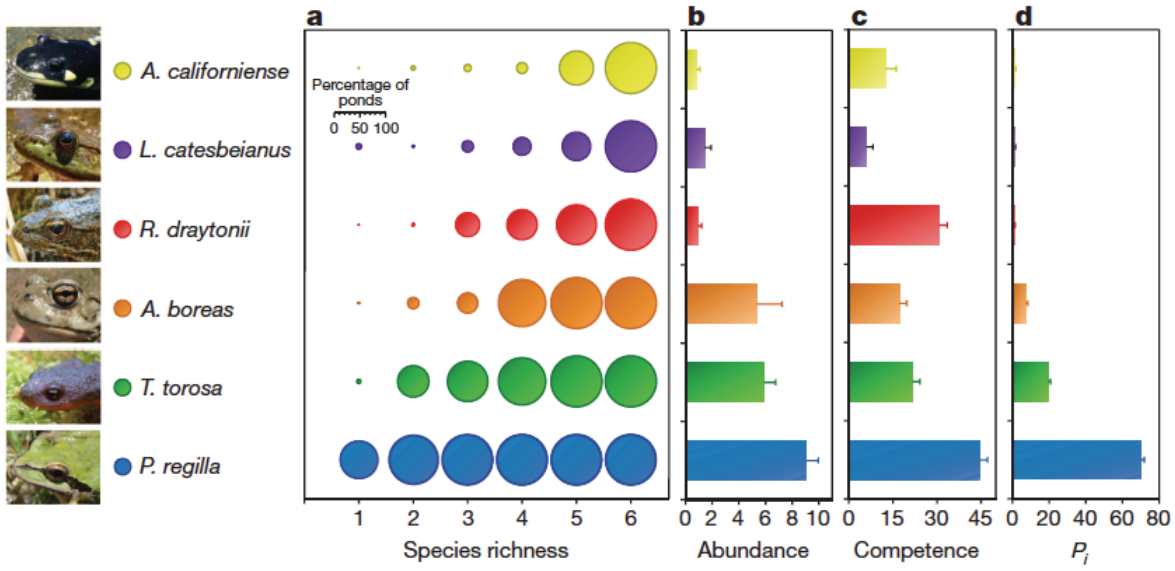


Figure 9: Relations observées entre la composition de communautés d'amphibiens et les traits fonctionnels des espèces incluses. a) La proportion des communautés incluant chaque espèce pour chaque niveau de diversité spécifique est indiquée par la taille des cercles. b) abondance moyenne de l'espèce dans les communautés où elle est présente, c) compétence de l'espèce pour le parasite *R. ondatrae* telle que mesurée en laboratoire, (d) indice de la contribution de l'espèce à la compétence globale de la méta-communauté (extrait de Johnson *et al.* 2013).

Une autre question essentielle est de savoir comment l'abondance des hôtes et des vecteurs varie avec la perturbation des habitats et les modifications de biodiversité (Johnson *et al.* 2015). Si l'augmentation de la richesse spécifique s'accompagne d'une augmentation de l'abondance des hôtes compétents, un effet d'amplification peut être observé, même si leur proportion dans la communauté est réduite. Les modèles épidémiologiques considérant l'augmentation du nombre d'espèces comme un phénomène additif concluent immanquablement à un effet d'amplification pour les maladies à transmission densité-dépendante (Dobson 2004; Mihaljevic *et al.* 2014). Il est pourtant évidemment nécessaire de prendre en compte les interactions interspécifiques telles que la prédation ou la compétition pour les ressources, qui maintiennent une corégulation des populations. Ces interactions permettent d'expliquer que certaines espèces surabondent dans des milieux ou que d'autres ne sont plus présentes, un phénomène connu sous le nom de « compensation de densité » par les écologues (Peres & Dolman; MacArthur *et al.* 1972; Henke & Bryant 1999; Michalski & Peres 2007). Ceci est d'autant plus important que la disparition d'espèces clés telles que les top-prédateurs constitue une signature typique des perturbations anthropogéniques (Estes *et al.* 2011). Il faut donc progresser vers une intégration des connaissances écologiques pour

prendre en compte les interactions trophiques et envisager des modifications de structures de communauté réalistes dans les études éco-épidémiologiques (Packer *et al.* 2003, Roche *et al.* 2012; Mihaljevic *et al.* 2014).

Finalement l'une des priorités est d'évoluer vers l'adoption d'une perspective multi-pathogènes. Chaque espèce vivante constitue un hôte pour une multitude de parasites. Ceux-ci représentent une part importante de la biodiversité dont la grande majorité reste encore à découvrir (Dobson *et al.* 2008; Poulin & Morand 2014). Si une diminution de la biodiversité peut favoriser la transmission de certains pathogènes portés par les hôtes persistants dans les communautés appauvries, elle entrainera vraisemblablement la coextinction de ceux qui dépendent des hôtes les moins résilients. Il a même été envisagé que les parasites puissent être les principales victimes de l'érosion de la biodiversité en cours (Dunn *et al.* 2009). Il existe une relation positive entre diversité d'hôtes vertébrés et diversité parasitaire sur de larges échelles géographiques (Dunn *et al.* 2010). Celle-ci s'explique simplement par une augmentation de l'hétérogénéité de l'habitat parasitaire (constitué ici par les communautés d'hôtes; Johnson *et al.* 2016), ou, autrement formulé, par l'idée que « la biodiversité entraine la biodiversité » (Johnson *et al.* 2015).

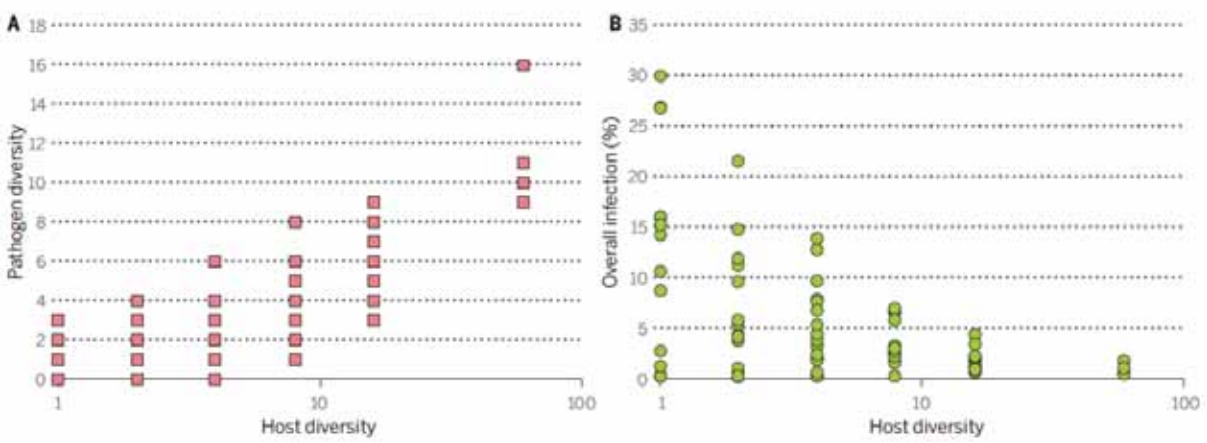


Figure 10: Effet de la diversité de plantes sur la prévalence de pathogènes fongiques dans des communautés manipulées expérimentalement. Les communautés incluant un plus grand nombre d'espèces de plantes abritent une plus grande diversité de pathogènes **(A)**, mais une prévalence totale plus faible **(B)** (extrait de Keesing & Ostfeld 2015, réalisé avec les données de Rottstock *et al.* 2014).

La diversité parasitaire n'équivaut toutefois pas au risque infectieux, et l'exploration simultanée de ces deux mesures est un objectif important. L'une des rares études menées en ce sens a permis de mettre en évidence (i) la relation positive attendue entre diversité de pathogènes et diversité de plantes, et (ii) un effet négatif de la diversité sur le pourcentage de plantes infectées et la sévérité des infections (Rottstock et al. 2014,

Figure **10**).

Cette ambivalence a également été observée dans une étude menée à l'échelle des pays en Asie du Sud-Est: Morand *et al.* (2014) montrent que le nombre de pathogènes répertoriés est corrélé positivement à la richesse spécifique de vertébrés, mais que la fréquence d'épidémies zoonotiques augmente avec le nombre d'espèces menacées de mammifères et d'oiseaux (Figure 11). Ces résultats font écho à l'idée que la diversité génétique contribue à limiter les épidémies de grande ampleur, bien qu'elle soit associée à un plus grand nombre d'épidémies mineures (Springbett *et al.* 2003). Dans le contexte actuel d'émergence de maladies zoonotiques, il est essentiel d'évaluer les conséquences de ces phénomènes sur le risque de passage de barrière d'espèce.

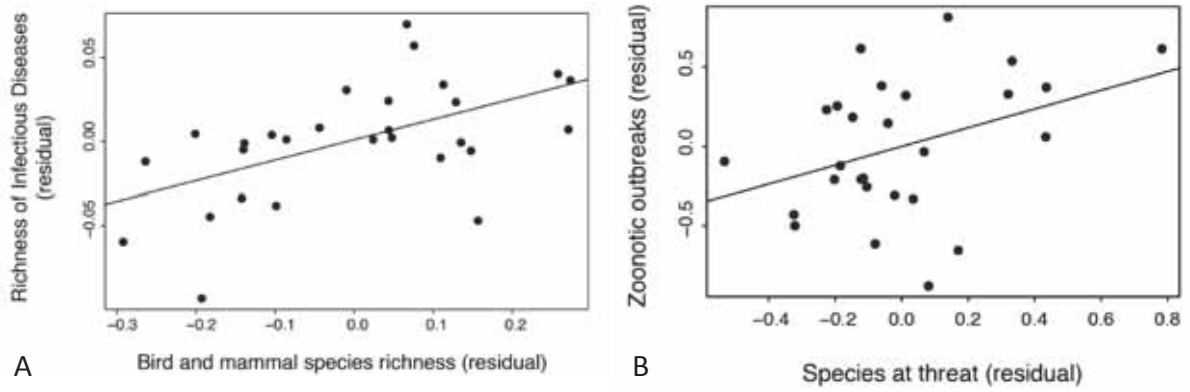


Figure 11: (A) Nombre d'agents pathogènes répertoriés en fonction de la richesse spécifique d'oiseaux et de mammifères dans différents pays d'Asie du Sud-Est. **(B)** Nombre d'épidémies zoonotiques répertoriées en fonction du nombre d'espèces menacées (extrait de Morand *et al.* 2014).

D. EFFET DE DILUTION ET MALADIES VECTORIELLES : RÔLE DU COMPARTIMENT VECTEUR

Bien que le concept d'effet de dilution se rapporte initialement au cas particulier des maladies vectorielles, et que ces dernières constituent une part importante des maladies infectieuses émergentes et réémergentes (Taylor *et al.* 2001), le rôle de la communauté d'arthropodes vecteurs n'a reçu que peu d'attention dans les recherches sur les liens existant entre biodiversité et maladies infectieuses. Si un certain nombre de travaux ont permis d'explorer les effets de la diversité et de la composition des communautés d'hôtes sur le cycle de transmission des pathogènes, le compartiment vectoriel n'est généralement pas considéré en détail.

Les arthropodes vecteurs représentent une composante à part entière des écosystèmes, et leur sensibilité aux perturbations d'origine humaine a vraisemblablement des conséquences épidémiologiques. L'une des premières questions soulevées est celle des effets de la diversité d'hôtes vertébrés sur l'abondance des vecteurs. Si une plus grande diversité d'hôtes

représente une plus grande quantité de ressources disponibles pour les arthropodes hématophages, il est envisageable qu'elle conduise à une augmentation de leur densité. Or, un effet de dilution induit par la présence d'un hôte non compétent peut être contrebalancé si elle entraine également une augmentation du nombre de vecteurs, *via* deux mécanismes interconnectés: (i) le ratio du nombre de vecteurs sur le nombre d'hôtes réservoirs est un facteur de transmission important et son augmentation conduit à une plus grande fréquence d'infection (Keeling & Rohani 2008), (ii) à fréquence d'infection donnée, une augmentation de la densité de vecteurs se traduit directement par une augmentation de la densité de vecteur infectés, et donc, du risque de transmission.

La relation entre abondance des vecteurs et diversité d'hôte est conditionnée par la capacité du vecteur à se nourrir sur différentes espèces d'hôtes et le caractère limitant de la prise de repas sanguin dans son cycle de vie. Ces conditions sont typiquement considérées comme réunies pour les tiques qui sont relativement généralistes et peu mobiles, dont la survie est dépendante du passage d'hôtes vertébrés à proximité. Sous réserve que l'augmentation de la diversité d'hôte s'accompagne d'une augmentation de leur densité totale, elle est susceptible de contribuer à la prolifération de la population de tiques. La non-prise en compte de ce phénomène dans la modélisation du risque de transmission peut conduire à un résultat biaisé, un problème qui fut au cœur des critiques de l'effet de dilution (Randolph & Dobson 2012). Il a été démontré empiriquement que l'extirpation de certaines espèces (*i.e.* un réduction de la biodiversité) peut conduire à une réduction drastique de la densité de tiques infectées et donc du risque de transmission (Laurenson *et al.* 2003; Rand *et al.* 2004; Swei *et al.* 2011).

Par opposition au cas des tiques, la densité de moustiques est classiquement considérée comme limitée par la disponibilité de sites de reproduction, et relativement indépendante de la densité d'hôtes dans le milieu (Dobson 2004). Ceci s'explique par leur grande mobilité qui permet une recherche active de repas sanguins. Toutefois, ici encore, des contre-exemples existent. Dans une étude menée au Pakistan, la présence de bétail à proximité du foyer a été associée à une prévalence accrue du paludisme, possiblement *via* un effet positif sur la population de moustiques, contredisant ainsi les prédictions du principe de zooprophylaxie analogue à l'effet de dilution.

Ces travaux focalisés sur un hôte vertébré supportant localement une proportion importante de repas-sanguins pour les vecteurs ne permettent pas de conclure à la généralité d'une relation positive entre diversité d'hôtes et prévalence des pathogènes. En revanche, ils constituent des contre-exemples à l'effet de dilution ayant un fort impact en termes de santé

publique. Il est donc essentiel de prendre en compte la démographie des vecteurs dans ce type d'étude.

La préférence trophique des vecteurs est un autre facteur conditionnant la réalisation d'un effet de dilution. Les arthropodes hématophages varient grandement dans leur degré de spécialisation sur certains hôtes. Les tiques du genre *Ixodes* sont par exemple typiquement considérées comme très généralistes, pouvant se nourrir sur une grande diversité de vertébrés incluant des oiseaux, des mammifères et des reptiles (James & Oliver 1990). A l'extrême inverse, les mouches hématophages de la famille des Nycteribiidae sont pour la plupart associées à une seule espèce de chauves-souris (Dick & Patterson 2007).

Un pathogène transmis par un vecteur très spécialisé devrait n'être que peu impacté par des changements de biodiversité, car la plupart des espèces présentes dans les communautés n'interviennent pas dans son cycle de transmission (Ostfeld & Keesing 2000b). Comme les cas de spécialisations extrêmes sont rares chez les vecteurs de maladies, une considération classique est que leurs préférences trophiques peuvent atténuer un effet de dilution mais ne devraient pas l'annuler totalement.

Une exploration plus détaillée peut toutefois conduire à des conclusions radicalement différentes. Dans une étude théorique, Miller & Huppert (2013) montrent que lorsque (i) la diversité d'hôtes n'est pas additive (i.e. que l'ajout d'un hôte entraine une diminution de la densité des autres), et que (ii) la densité des vecteurs ne dépend pas de la densité d'un hôte en particulier, l'ajout d'un hôte non-compétent peut entrainer une amplification du cycle de transmission dès lors que le vecteur affiche une préférence trophique (même modérée) pour l'hôte le plus compétent. La présence d'une espèce non-compétente peut provoquer une augmentation de la fréquence des interactions vecteur-hôte si l'abondance de ce dernier diminue mais que les vecteurs, toujours aussi abondants, ne sont pas déviés proportionnellement sur l'hôte non-compétent. Ce mécanisme peut mener à une relation non-monotone entre la proportion de l'hôte non-compétent dans la communauté et le taux de reproduction de base du pathogène qui, selon les auteurs, pourrait expliquer l'observation de patrons contradictoires (Figure 12). Leurs résultats démontrent que l'effet de dilution ne peut s'appliquer que dans un espace réduit de paramètres incluant la préférence trophique des vecteurs, la variabilité de compétence et l'abondance relative des hôtes. Cela rend la prédiction d'une relation négative entre biodiversité et risque de transmission beaucoup plus complexe que si elle était simplement dépendante de quelques conditions qualitatives (Encadré 1).

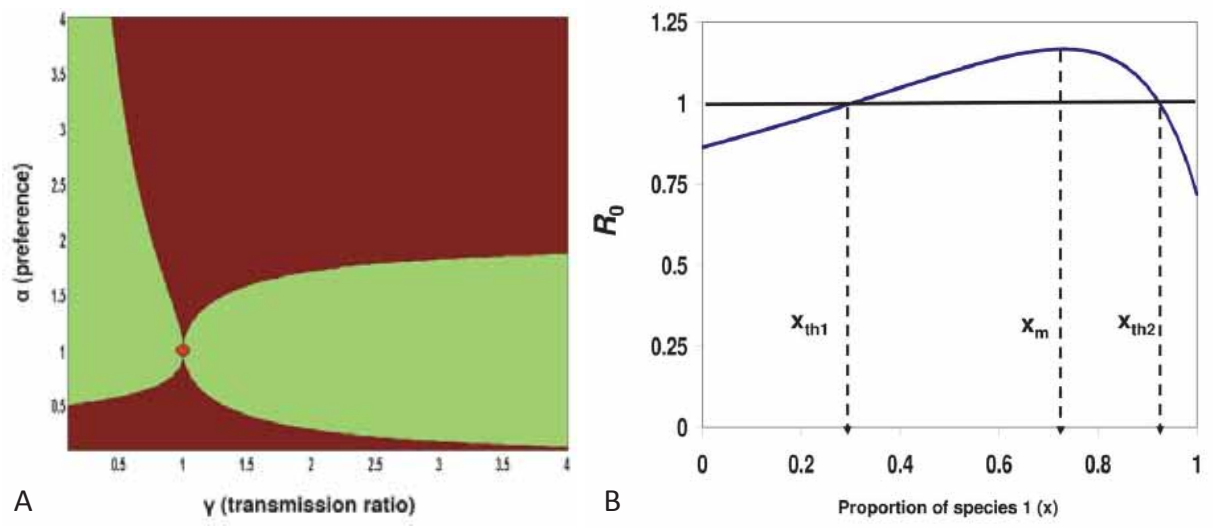


Figure 12: Modélisation du taux de reproduction de base (R0) pour un système comprenant deux espèces hôtes avec une abondance totale constante et une espèce de vecteur. **(A)** Comportement qualitatif du R0 en fonction de la préférence trophique du vecteur (α=préférence pour l'espèce hôte 2/espèce hôte 1) et du ratio de compétence entre les espèces hôtes (γ=compétence de l'espèce hôte 2/compétence de l'espèce hôte 1). Dans la zone verte, R0 diminue lorsque la proportion de l'hôte le moins compétent augmente (effet de dilution). Dans la zone brune, R0 augmente (effet d'amplification) puis diminue lorsque la proportion de l'hôte le moins compétent augmente. Au niveau du point rouge (α= γ=1), R0 est constant quelque soit la proportion des deux espèces. **(B)** Relation entre le R0 et la proportion de l'hôte le moins compétent dans la communauté avec α=3.3 et γ=1.2, i.e. le vecteur a une préférence trophique marquée pour un hôte légèrement plus compétent que l'autre (extrait de Miller & Huppert 2013).

Par ailleurs, les arthropodes vecteurs, au même titre que les hôtes vertébrés, appartiennent à un écosystème avec lequel ils interagissent. De plus, beaucoup de pathogènes peuvent être transmis par plusieurs espèces de vecteurs, souvent proches phylogénétiquement, mais pouvant avoir des caractéristiques fonctionnelles différentes en regard de la transmission. Il est donc important de considérer les modifications de la communauté d'arthropodes vecteurs de manière globale et dans un cadre écologiquement réaliste. Roche *et al.* (2013) proposent un modèle épidémiologique dans lequel la structure des communautés d'hôtes et de vecteurs est déterminée par des lois empiriques de distribution d'abondance établies de longue date en écologie (Fisher *et al.* 1943; Preston 1948). Leurs résultats indiquent que si une augmentation de la variabilité de compétence et de la diversité (indice de Shannon) des hôtes peut conduire à un effet de dilution, une augmentation de la variabilité de compétence des vecteurs est relativement sans effet (pour une compétence moyenne équivalente), tandis qu'une augmentation de leur diversité a tendance à amplifier le risque d'épidémie et la prévalence maximale atteinte. Ce dernier résultat découle du fait que le modèle implique une corrélation positive entre diversité et abondance des vecteurs. Ainsi, même lorsque des espèces vectrices contribuant faiblement à la transmission sont ajoutées à la communauté, le ratio vecteurs-hôtes global augmente.

D'autre part, une augmentation de la diversité fonctionnelle des vecteurs peut suffire à amplifier la transmission: Park *et al.* (2015) montrent que la prévalence de la maladie hémorragique transmise par différentes espèces de moucherons du genre *Culicoides* aux Etats-Unis augmente avec la diversité de ces derniers, de par la complémentarité de leur période d'activité qui résulte en une fenêtre temporelle de transmission plus large.

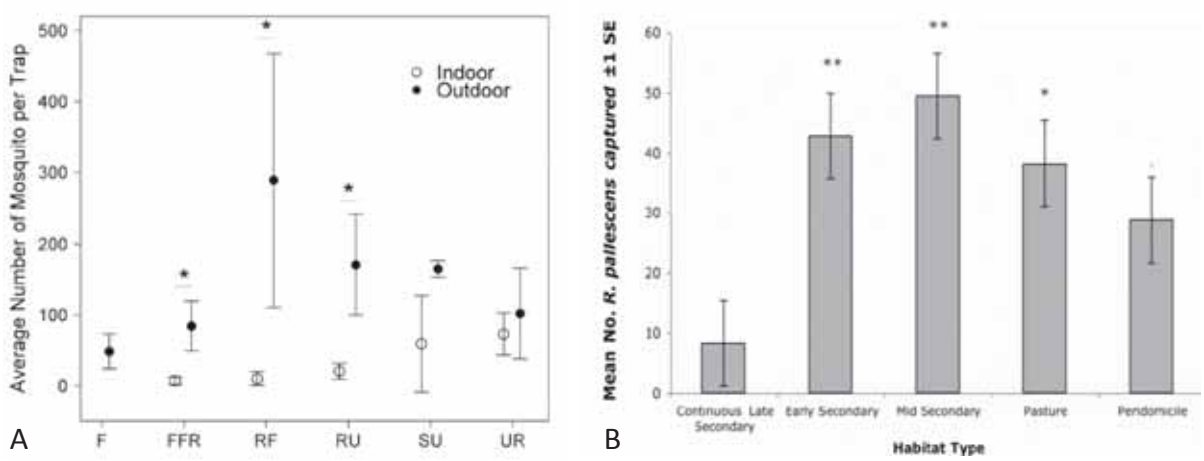


Figure 13: (A) Nombre moyen de moustiques collectés par piège-nuit en Thaïlande suivant le milieu considéré (F=forêt, FFR=forêt fragmentée, RF=champ de riz, RU=rural, SU=périurbain, UR=urbain; extrait de Thongsripong *et al.* (2013). **(B)** Nombre moyen de punaises (*Rhodnius pallescens,* vecteur de la maladie de Chagas) piégées par site le long d'un gradient d'anthropisation (extrait de Gottdenker *et al.* 2011).

Les questions déjà évoquées plus haut, concernant les règles d'assemblage des communautés en regard de la transmission doivent être explorées chez les arthropodes vecteurs. Plusieurs études tendent à montrer que les arthropodes vecteurs de pathogènes surabondent dans les milieux perturbés par l'homme (Friggens & Beier 2010; Gottdenker *et al.* 2011; Thongsripong *et al.* 2013; Figure 13), et que ceux-ci peuvent être régulés par la présence de prédateurs et de compétiteurs (Carlson *et al.* 2009). Toutefois, l'existence de contre-exemples (Walsh *et al.* 1993) et le caractère anthropocentré de ces recherches permettent difficilement de conclure à la généralité du phénomène. Il est nécessaire d'étudier plus globalement le comportement des communautés d'arthropodes hématophages face aux perturbations anthropogéniques, en particulier suite aux modifications des communautés d'hôtes vertébrés dont ils dépendent.

3. LES LEISHMANIOSES : ASPECTS CLINIQUES ET BIOLOGIQUES

A. GÉNÉRALITÉS

Les leishmanioses sont un ensemble de maladies causées par des parasites du genre *Leishmania* Ross, 1903. Les *Leishmania* sont des protozoaires flagellés de la famille des Trypanosomatidae (Kinetoplastida), transmis par la piqûre de petits diptères hématophages, les phlébotomes (Psychodidae, Phlebotominae). La plupart des leishmanioses affectant l'homme sont zoonotiques, leur cycle de transmission impliquant des hôtes réservoirs mammifères domestiques et sauvages. Ces maladies se manifestent par différentes formes cliniques chez l'homme, dépendant en grande partie de l'espèce de *Leishmania* en cause: leishmaniose viscérale (LV), la leishmaniose cutanée localisée (LCL) ou cutanée diffuse (LCD) et la leishmaniose cutanéomuqueuse (LCM).

L'homme est confronté aux leishmanioses depuis des temps anciens, comme l'indique l'existence de descriptions de lésions évoquant fortement la leishmaniose cutanée datant d'environ 2500 ans av. J.-C., et la découverte d'ADN leishmanien chez des momies de l'Egypte antique (Zink *et al.* 2006; Akhoundi *et al.* 2016). Les premières descriptions cliniques détaillées de la leishmaniose viscérale datent quand à elles du 18ème siècle en Inde, mais le parasite ne fut formellement identifié qu'au début du 20ème siècle. Aujourd'hui, les leishmanioses sont largement réparties dans le monde, principalement en zone intertropicale, mais débordant en zone tempérée (Figure 14). Elles sont endémiques dans 98 pays appartenant à quatre continents. On estime qu'elles sont responsables de plus d'un million de cas chaque année, et de 20,000 à 40,000 décès (Alvar *et al.* 2012). Touchant essentiellement les populations les plus pauvres, elles sont classées par l'Organisation Mondiale de la Santé (OMS) comme maladies tropicales négligées.

B. CYCLE PARASITAIRE, PHYSIOPATHOLOGIE ET FORMES CLINIQUES

D'après Dedet (2009).

Les *Leishmania*, présentes sous forme amastigote dans la peau ou le sang d'un mammifère infecté, sont ingérées par un phlébotome à l'occasion d'un repas sanguin (Figure 15). Dans le tube digestif du vecteur, ces parasites se transforment rapidement en promastigotes puis traversent la membrane péritrophique. Après une phase de multiplication

dans la lumière de l'intestin, ils migrent vers la partie antérieure du tube digestif où ils deviennent infestant (les promastigotes métacycliques), qui seront injectés chez un nouvel hôte mammifère lors d'une piqûre ultérieure.

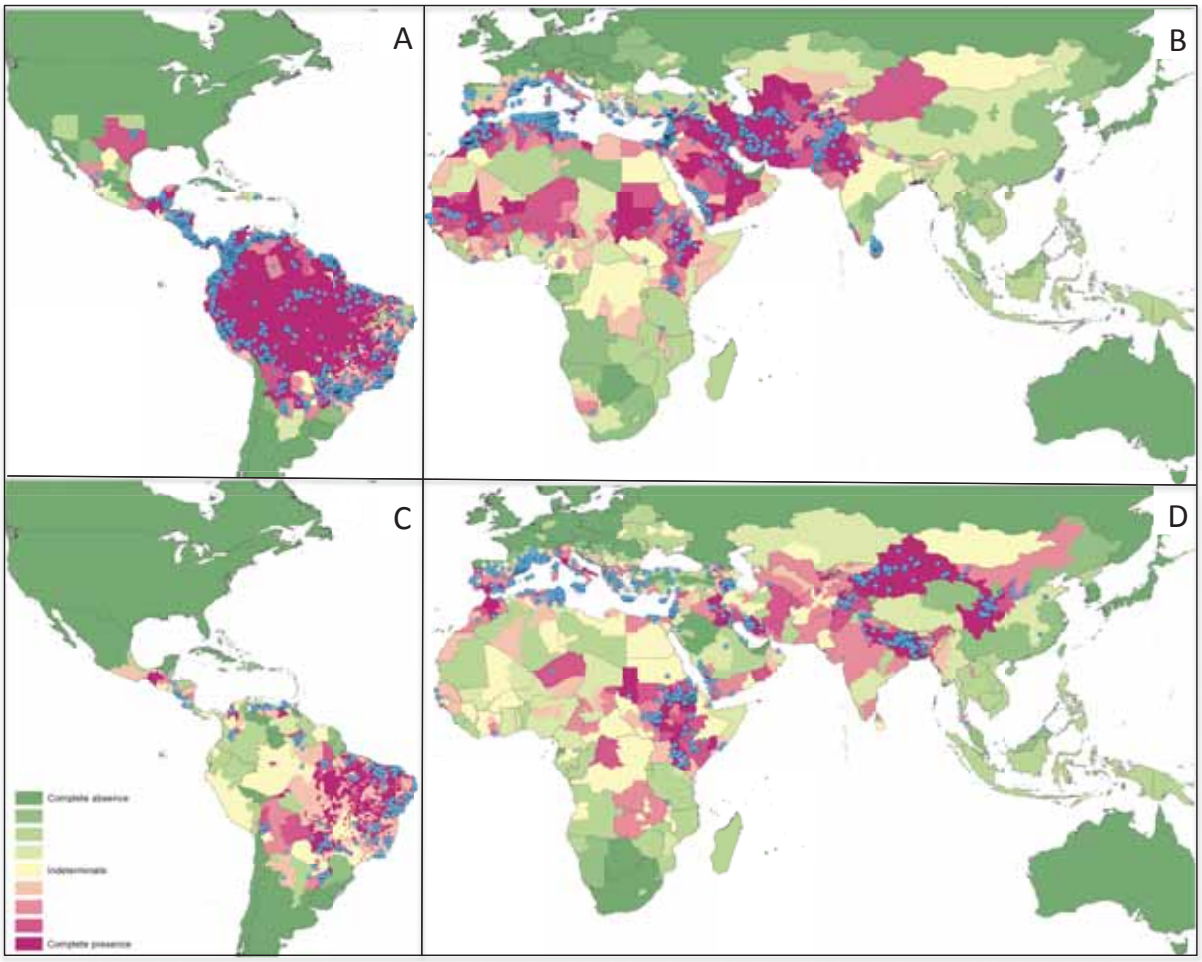


Figure 14: Niveau de consensus pour la présence des leishmanioses cutanées (A et B) et viscérales (C et D) dans le Nouveau (A et C) et l'Ancien Monde (B et D), basé sur la confrontation de sources de données variées. Les points bleus indiquent des points ou des centroïdes d'occurrence (Pigott *et al.* 2014).

Dans le derme, l'installation des parasites est favorisée par des molécules contenues dans la salive du phlébotome ayant une action immunosuppressive locale (Kamhawi 2000). Les promastigotes métacycliques sont phagocytés par des macrophages du derme, et intègrent une vacuole parasitophore dans laquelle ils repassent en forme amastigote. Les parasites résistent à la digestion par les macrophages grâce à différents mécanismes de sabotage moléculaire (Arango Duque & Descoteaux 2015). Après leur multiplication et l'éclatement du macrophage, les *Leishmania* peuvent infecter de nouvelles cellules phagocytaires localement, ou migrer vers d'autres tissus, en fonction du tropisme de l'espèce en cause et les caractéristiques de l'hôte. Les défenses immunitaires permettent généralement de limiter le

développement du parasite, donnant lieu en général à une infection asymptomatique (Costa *et al.* 2002; Bañuls *et al.* 2011). Dans de plus rares cas, l'infection se manifeste cliniquement.

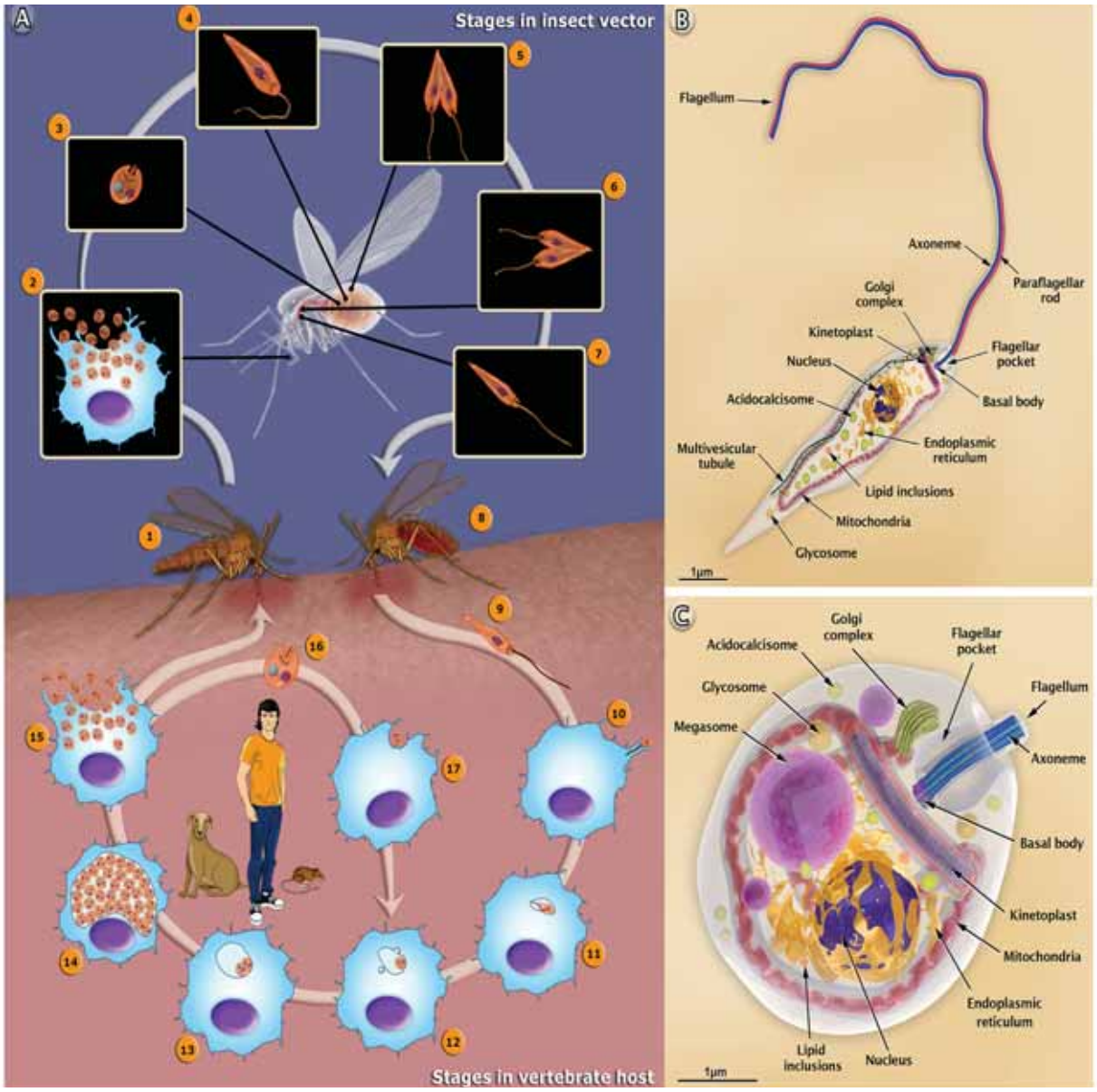


Figure 15: (A) Cycle de vie des *Leishmania*: (1) La femelle de phlébotome prend un repas sanguin sur un hôte infecté, (2,3) elle ingère des macrophages infectés par des formes amastigotes qui sont libérées après explosion de la cellule, (4) les amastigotes se transforment en promastigotes procycliques, (5) multiplication dans la lumière de l'intestin du phlébotome, (6) migration vers la partie antérieure du tube digestif (valve stomodéale) et nouvelle phase de multiplication, (7) les promastigotes se transforment en promastigotes métacycliques (formes infestantes), (8,9) le phlébotome injecte les promastigotes métacycliques chez un nouvel hôte par régurgitation lors d'un nouveau repas sanguin, (10) les promastigotes métacycliques infectent des macrophages, (11) transformation en forme amastigote, (12) les amastigotes s'attachent à la paroi des vacuoles parasitophores, (13,14) multiplication des amastigotes dans la vacuole, (15) les amastigotes sont libérés après explosion cellulaire, (16,17) les amastigotes libérés peuvent infecter de nouveaux macrophages. **(B)** Organisation structurale du promastigote. **(C)** Organisation structurale des amastigotes (extrait de Teixeira *et al.* 2013).

La plupart des espèces de *Leishmania* ont un tropisme cutané exclusif. La leishmaniose cutanée localisée est la forme la plus commune et résulte de la multiplication du parasite au site d'inoculation (y compris pour les *Leishmania* viscérotropes). Après une phase

d'incubation allant d'un à quatre mois, une lésion apparaît consistant d'abord en une petite papule inflammatoire. Celle-ci grossit ensuite de manière constante, et évolue le plus souvent vers une forme ulcérée dite « humide », nettement circonscrite par un bourrelet périphérique et pouvant s'élargir jusqu'à une dizaine de centimètres. D'autres formes « sèches » ou tuberculoïdes peuvent également être observées. Ces lésions ne sont théoriquement pas douloureuses sauf en cas de surinfection et finissent par guérir spontanément, mais ceci peut prendre plusieurs mois, voire plusieurs années, et laissent des cicatrices indélébiles pouvant être particulièrement invalidantes lorsqu'elles se trouvent au niveau du visage. Le plus souvent chez des sujets anergiques ou atteints de déficience immunitaire, il peut être observé des formes cutanées diffuses. Celles-ci sont généralement causées par *Leishmania aethiopica* dans l'ancien monde et par *L. amazonensis* dans le nouveau monde. Ces formes sont caractérisées par l'apparition de nodules un peu partout sur le corps, qui grossissent jusqu'à se rejoindre pour former de larges plaques, évoquant des lésions lépreuses. La leishmaniose cutanée est causée par une quinzaine d'espèces de *Leishmania* différentes depuis l'Amérique du Sud jusqu'en Asie centrale (Figure 14).

Chez les sujets sensibles, les deux espèces de *Leishmania* viscérotropes (*L. donovani* et *L. infatum*) disséminent à tous les organes du système des phagocytes mononucléés, causant une leishmaniose viscérale. Après une période d'incubation de trois à six mois, la phase d'état évolue de manière chronique, avec aggravation progressive des symptômes : fièvre intermittente, anémie, amaigrissement et hyperplasie des organes atteints: rate, foie et éventuellement ganglions lymphatiques. Plus tard, des complications digestives, pulmonaires et hémorragiques peuvent survenir. L'anémie est l'un des signes les plus caractéristiques : en Inde, la maladie porte le nom de kala-azar (peste noire), en référence à l'aspect gris terreux que prend la peau des patients. A un stade avancé, la splénomégalie associée à un fort amaigrissement peut également être frappante (Figure 16). Sans traitement, la maladie évolue quasi-systématiquement vers la mort. Si un traitement est mis en place suffisamment tôt, le pronostic est favorable, mais les parasites ne sont pas forcément tous éliminés. Certains peuvent se maintenir dans l'organisme ou donner lieu à une complication dermique particulière, particulièrement pour l'espèce *L. donovani* : la leishmaniose cutanée post-kala-azar. La leishmaniose viscérale connait une distribution géographique très large, mais 90% des cas sont concentrés dans 6 pays selon l'OMS (Alvar *et al.* 2012): Inde, Bangladesh, Soudan, Soudan du Sud, Ethiopie et Brésil.

Figure 16: Sujets atteints de **(A)** leishmaniose viscérale, **(B)** leishmaniose cutanée post-kala-azar, **(C)** leishmaniose cutanée localisée et **(D)** leishmaniose cutanéo-muqueuse (source= OMS).

Enfin, certaines espèces et en particulier *L. braziliensis* mais aussi l'espèce *L. guyanensis* sont responsables d'une forme clinique distincte : la leishmaniose cuntanéomuqueuse. Cette forme évolue en deux temps : l'atteinte initiale se manifeste comme une LCL classique, mais peut être suivie par une atteinte muqueuse, parfois plusieurs années après la guérison de la première lésion. Celle-ci commence généralement au niveau de la muqueuse nasale, et peut évoluer vers une destruction de la cloison nasale et une atteinte sévère de la muqueuse buccale, du palais et du pharynx. L'atteinte muqueuse peut progresser jusqu'à la mise en communication les fosses nasales et la cavité buccale, et conduire à de la dysphagie ou de la dysphonie. Les lésions des stades avancés sont particulièrement mutilantes et peuvent avoir un impact physiologique et psychosociologique sévère sur le sujet. *L. braziliensis* n'est présent que dans le nouveau monde, du sud du Mexique au nord de l'Argentine.

Il faut noter cependant que, même si l'on observe globalement une association entre espèces de *Leishmania* et formes cliniques chez l'homme, chaque espèce pathogène pour l'homme peut provoquer une grande diversité de symptômes, ce qui rend le diagnostic spécifique difficile (Bañuls *et al.* 2011).

C. PRÉVENTION, DIAGNOSTIC ET TRAITEMENT

D'après Dedet (2009) et Buffet *et al.* (2011)

Bien que des vaccins procurant une protection partielle pour le chien aient été mis sur le marché au Brésil et en Europe (Leish-Tec®, Hertape; Leishmune®, Fort Dodge; Canileish®, Virbac; Gradoni 2015), et que plusieurs candidats soient en cours de développement pour l'homme (Gillespie *et al.* 2016), aucun n'est encore disponible. La prévention des leishmanioses repose donc sur l'évitement des piqûres de phlébotomes, par l'emploi d'insecticides ou de répulsifs et de moustiquaires imprégnées. L'établissement d'un programme de lutte globale est compliqué du fait de la diversité de structure des foyers d'infection (voir partie 5.A).

Le diagnostic de la maladie se fait par la recherche du parasite, de son ADN, ou par la mise en évidence immunologique de l'infection. Pour les méthodes parasitologiques directes, le prélèvement se fait par curetage ou biopsie de la lésion pour la LC ou la LCM, et principalement par ponction de moelle osseuse ou prélèvement de sang pour la LV. La mise en évidence du parasite peut se faire sur frottis coloré ou par mise en culture du prélèvement. Cette dernière méthode, bien que plus fastidieuse, dispose d'une meilleure sensibilité et permet de conserver la souche pour des études ultérieures.

Le diagnostic moléculaire, par PCR en particulier, s'est beaucoup développé du fait des multiples avantages qu'il présente: grande sensibilité et spécificité, rapidité, et possibilité d'accéder à l'identification de l'espèce de *Leishmania* en cause (Marques *et al.* 2006; Ruiter *et al.* 2014). Il nécessite par contre des infrastructures, du matériel et du personnel spécialisé qui ne sont pas toujours disponibles dans les régions endémiques. Des alternatives reposant sur des techniques d'amplification isothermales, plus facilement applicables sur le terrain, ont été développées (Saldarriaga *et al.* 2016; Nzelu *et al.* 2016).

Des tests immunologiques sont également employés. La LV et la LCD entrainent généralement la production d'anticorps circulants qui peuvent être détectés par immunofluorescence indirecte (IFI), méthode immuno-enzymatique (ELISA) ou agglutination directe (DAT) par exemple. Ces techniques présentent l'avantage d'être non-invasives et relativement faciles à mettre en œuvre (des kits commerciaux existent). Elles sont utiles en

première intention mais souffrent d'un manque de spécificité du fait de possibles réactions croisées. De plus, la présence d'anticorps peut être associée à une infection asymptomatique ou passée, et n'indique pas nécessairement l'établissement de la maladie au moment du test. La LCL et la LCM s'accompagnent plutôt d'une réaction immunitaire de type cellulaire pouvant être mise en évidence par un test d'hypersensibilité retardée (test Montenegro), qui constitue en l'observation d'une réaction cutanée suite à l'injection intradermique d'antigènes parasitaires. Ce test a été mis au point il y a près d'un siècle au Brésil (Montenegro 1926) et représente encore l'outil diagnostique principal dans certaines régions (Antonio *et al.* 2014). Il bénéficie d'un bon niveau de sensibilité et de spécificité, et il est facile d'utilisation. Cependant, son résultat n'est pas immédiat et sa positivité persiste longtemps après la guérison.

Le traitement des leishmanioses est compliqué par le coût et la toxicité des produits disponibles, la sensibilité variable des différentes espèces de *Leishmania* selon les molécules utilisées et l'existence de résistances chez certaines souches (Buffet *et al.* 2011). Les médicaments anti-leishmaniens classiques comprennent:

• les sels pentavalents d'antimoine (Glucantime® et Pentostam®) qui agissent par inhibition de la synthèse d'adénosine triphosphate (ATP)

• l'amphotéricine B (et ses formulations lipidiques, Fungizone®, AmBisome®), un antifongique puissant de la famille des polyènes, qui provoque une altération de la membrane cellulaire des parasites

• la pentamidine (Pentacarinat®), une diamine aromatique agissant par inhibition de la synthèse d'ADN.

L'utilisation de ces molécules est limitée par des effets secondaires importants et des voies d'administration et des posologies contraignantes (*e.g.* voie intraveineuse lente pendant plusieurs jours pour l'amphotéricine B). Les dérivés d'antimoine ont longtemps été les plus utilisés et restent encore aujourd'hui employés en première intention dans certains cas de LC (Buffet *et al.* 2011; Tableau 2). L'apparition de souches résistantes aux dérivés de l'antimoine, l'épidémie de co-infection *Leishmania*-VIH, et la meilleure tolérance des formulations lipidiques de l'amphotéricine B font que ces dernières sont de plus en plus utilisées dans le monde entier, et constituent d'ores et déjà le traitement de première intention pour la LV (Alvar *et al.* 2008; Boer *et al.* 2009). La pentamidine n'est efficace qu'à forte dose dans le traitement de la LV, pour lequel elle n'est plus préconisée, mais son usage reste recommandé dans certaines formes de LC. Plus récemment, un anti-leishmanien oral et bien toléré, la miltéfosine (Impavido®), a été employé avec succès pour le traitement de la LV en

Inde. Toutefois, de récents cas d'échecs ont été rapportés, et son efficacité semble plus variable dans d'autres zones géographiques ou pour le traitement de la LC (Monge-Maillo & López-Vélez 2015). D'autres molécules ont été proposées et font l'objet d'essais thérapeutiques, mais ne sont pas encore utilisées en pratique courante.

Tableau 2: Synthèse du référentiel proposé pour la prise en charge des principales formes de Leishmanioses rencontrées en France (Buffet *et al.* 2011).

Forme clinique/Espèce de *Leishmania*	**Traitement recommandé**
LV	Amphotéricine B
LC à *L. major*	1. Abstention si lésion peu gênante et patient consentant 2. Antimoniate de méglumine intralésionnel + cryothérapie 3. Si plus de 4 lésions, ou localisation des lésions incompatible avec traitement local, ou échec de traitement 2, amphotéricine B
LC à *L. tropica ou L. infantum*	1. Antimoniate de méglumine intralésionnel + cryothérapie 2. Si plus de 4 lésions, ou localisation des lésions incompatible avec traitement local, ou échec de traitement 2, amphotéricine B
LC à *L. guyanensis ou L. panamensis*	1. Iséthionate de pentamidine par voie intramusculaire 2. Si deux échecs du traitement 2, antimoniate de méglumine par voie intraveineuse lente ou IM, ou miltéfosine
LC à *L. braziliensis*	1. Antimoniate de méglumine par voie IV lente ou IM 2. Si deux échecs du traitement 2, amphotéricine B ou miltéfosine
LM à *L. braziliensis*	1. Antimoniate de méglumine par voie IV lente ou IM + pentoxyfilline 2. Si échec du traitement 2, amphotéricine B et/ou miltéfosine

4. LES *LEISHMANIA*

D'après Bañuls *et al.* (2007), Cantacessi *et al.* (2015) et Akhoundi *et al.* 2016)

A. ASPECTS MORPHOLOGIQUES ET GÉNOMIQUES

Les *Leishmania* sont, comme tous les kinétoplastidés, caractérisés par une mitochondrie unique comprenant un génome exceptionnellement large et condensé, visible en microscopie par les techniques de coloration classiques : le kinétoplaste. La morphologie globale des *Leishmania* varie ensuite selon le stade parasitaire. Le stade promastigote chez le vecteur, est une forme extracellulaire mobile caractérisée par une forme allongée, une taille variant entre 10 et 25 µm, la présence d'un flagelle libre en position antérieure et une position du kinétoplaste à la base du flagelle (Figure 17). Le stade amastigote, chez l'hôte vertébré, est une forme intracellulaire immobile caractérisée par une forme ovoïde, un diamètre de 2 à 6 µm, un flagelle de courte taille inclus dans un repli cellulaire et un kinétoplaste en position juxtanucléaire. Les parasites se multiplient par divisions binaires simples au cours des deux stades.

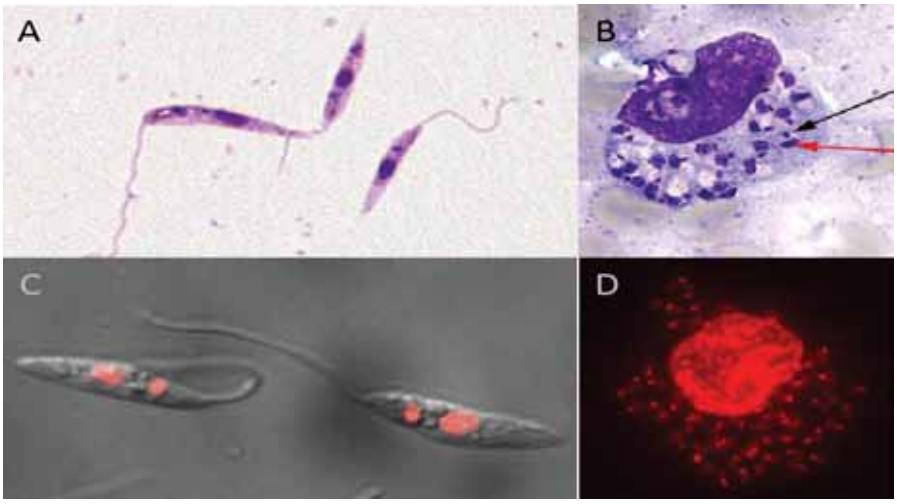


Figure 17: (A) Formes promastigotes de *Leishmania* sp. en microscopie après coloration classique (source: CDC). **(B)** Formes amastigotes de *Leishmania* sp. dans un macrophage infecté. Le noyau (flèche noire) et le kinétoplaste (flèche rouge) sont distinguables (source: CDC). **(C)** Formes promastigotes de *L. infantum* en microscopie à contraste interférentiel après marquage fluorescent de l'ADN (nucléaire et kinétoplastique) au DAPI (cliché : Baptiste Vergnes). **(D)** Amastigotes intracellulaires de *L. infantum* après marquage fluorescent de l'ADN au DAPI montrant le noyau du macrophage et l'ADN (nucléaire et kinétoplastique) des parasites (cliché : Baptiste Vergnes).

Le génome leishmanien est constitué du génome nucléaire et du génome kinétoplastique. Plus de dix ans après le lancement du *Leishmania Genome* Network, le premier génome nucléaire de *Leishmania* séquencé fut celui de *L. major* (Ivens *et al.* 2005),

suivi rapidement par celui de *L. braziliensis* et de *L. infantum* (Peacock *et al.* 2007). Le développement des technologies de séquençage haut-débit a ensuite permis la production de génomes de référence pour *L. mexicana* (Rogers *et al.* 2011), *L. donovani* (Downing *et al.* 2011), *L. amazonensis* (Real *et al.* 2013), *L. panamensis* (Llanes *et al.* 2015), tandis que d'autres sont en cours. Le génome leishmanien a une taille variant entre 29 et 33 Mb selon les espèces, et est constitué de 34 à 35 chromosomes.

Bien qu'ils soient classiquement considérés comme des organismes diploïdes, une grande plasticité génomique est observée chez les *Leishmania.* Chaque chromosome peut être monosémique, disomique ou trisomique selon la cellule considérée au sein d'une même souche (Sterkers *et al.* 2012). Cette « aneuploïdie mosaïque » constitue probablement une stratégie d'adaptation puissante pour le parasite. Le génome leishmanien est également caractérisé par des taux de GC et de séquence codante élevés, la répétition de gènes codants en tandem, la présence de longs groupes polycistroniques et la quasi-absence d'introns (Uliana *et al.* 2008). La régulation de l'expression génique ne se fait pas au niveau de la transcription chez les *Leishmania*, mais semble reposer sur la plasticité du nombre de chromosomes et de copies de gènes, ainsi que sur des mécanismes de maturation et de dégradation des ARN messagers.

L'ADN kinétoplastique constitue le génome de l'unique mitochondrie des kinétoplastidés et représente 10 à 20% de l'ADN total de la cellule (Simpson 1987). Il s'agit d'un réseau de molécules circulaires concaténées de deux types: les maxicercles et les minicercles (Figure 18). Les maxicercles sont des molécules d'environ 20 kb présentes en quelques dizaines de copies homogènes et contiennent, comme chez d'autres eucaryotes supérieurs, les gènes des ARN ribosomaux de la mitochondrie et des sous-unités protéiques de la chaine respiratoire. Les minicercles sont de courtes molécules d'environ 1 kb présentes en des dizaines de milliers de copies. Ils codent pour des ARN guides intervenant dans le mécanisme d'édition des ARN messagers du maxicercle. Celui-ci consiste en la délétion ou l'insertion d'uridine au niveau de sites internes des ARN messagers pré-édités (Read *et al.* 2015) et est nécessaire à leur traduction en protéines. Différentes « classes » de minicercles coexistent dans le génome, ciblant des sites d'édition distincts.

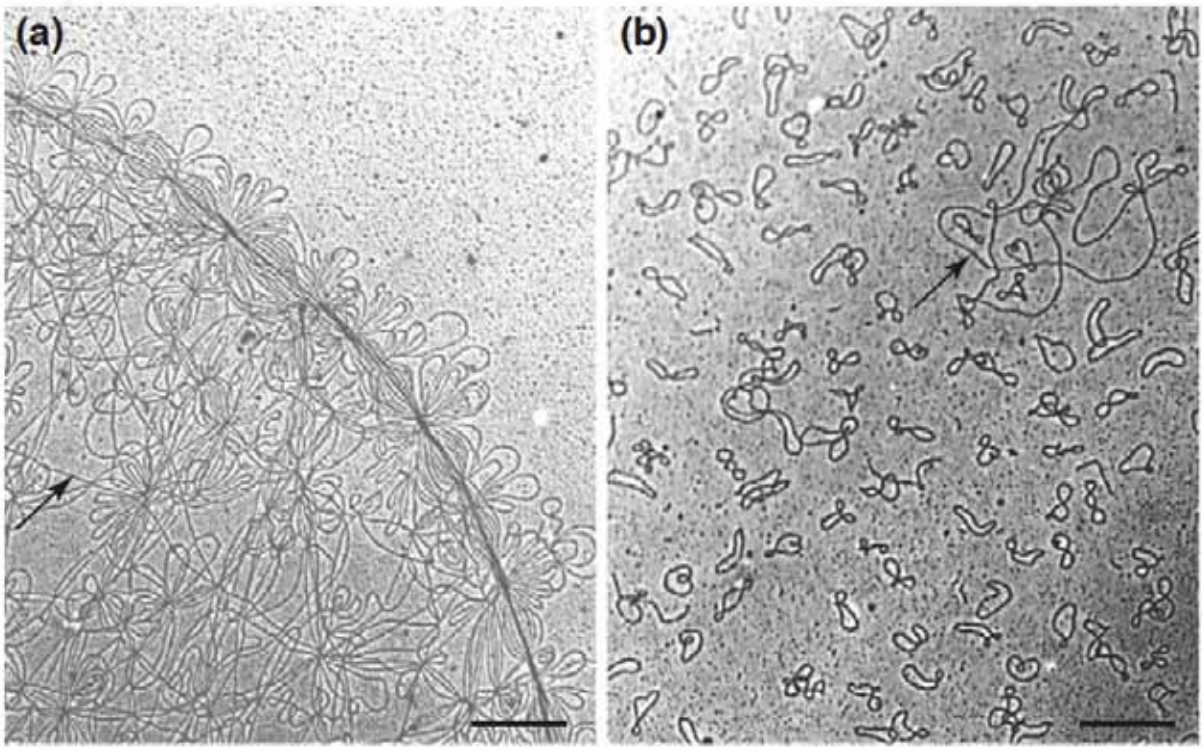


Figure 18: ADN kinétoplastique de *Crithidia fasciculata* (Trypanosomatidae) en microscopie électronique: **(a)** réseau d'ADN kinétoplastique concaténé intact. Les petites boucles sont des minicercles, et les brins plus longs sont des morceaux de maxicercles (flèche; barre d'échelle=0.3 μm); **(b)** ADN kinétoplastique décaténé par la topoisomérase II, permettant de visualiser des minicercles isolés et un maxicercle (flèche, barre d'échelle=0.5 μm) (extrait de Shapiro *et al.* 1999).

On considère actuellement que les *Leishmania* ont un mode de reproduction mixte avec une alternance entre clonalité et reproduction sexuée de types endogamique (recombinaison sexuée entre individus apparentés) et allogamique (recombinaison sexuée entre individus non-apparentés) (Rougeron *et al.* 2015). Cette alternance de mode de reproduction, assez classique chez de nombreux microorganismes, a été démontrée de manière empirique par des études de génétique des populations et validée par des études expérimentales chez le vecteur (Akopyants *et al.* 2009; Romano *et al.* 2014). Il semble, cependant, que selon l'écosystème et les espèces de *Leishmania*, ces parasites font plus ou moins appel à l'un ou l'autre de ces deux modes (Rougeron *et al.* 2015).

B. ASPECTS TAXONOMIQUES ET IDENTIFICATION

Les *Leishmania* appartiennent à la famille des Trypanosomatidae, qui contient également les trypanosomes responsables de la maladie du sommeil en Afrique intertropicale (*Trypanosoma brucei*) et de la maladie de Chagas en Amérique (*T. cruzi*), les pathogènes végétaux du genre *Phytomonas*, ainsi qu'un grand nombre d'autres parasites exclusifs d'insectes.

Les différentes espèces de *Leishmania* ne sont pas bien distinguables morphologiquement, et leur classification a fait l'objet de nombreux travaux, se basant dans la première partie du 20ème siècle sur une combinaison de critères éco-biologiques tels que la distribution géographique, la manifestation clinique ou l'espèce de phlébotome vectrice, ainsi que sur des caractères culturaux et immunologiques. Une étape importante de la classification des *Leishmania* a été la distinction du genre *Sauroleishmania*, incluant toutes les espèces infectant les lézards (Killick-Kendrick *et al.* 1986), et la séparation du genre *Leishmania* en deux sous-genres sur la base de leur localisation dans le vecteur: le sous-genre *Leishmania* (anciennement Suprapylaria) se développant uniquement en position antérieure du pylore et le sous-genre *Viannia* (anciennement Peripylaria) se développant au niveau du pylore et en position postérieure (Lainson & Shaw 1987).

Le développement de méthodes moléculaires, depuis les années 1970, a permis beaucoup de progrès dans la classification des *Leishmania*. En particulier, la technique isoenzymatique, qui met en évidence le polymorphisme enzymatique par les différences de mobilités électrophorétiques, reflet de la variabilité allélique des gènes qui les codent, est aujourd'hui encore considérée comme une méthode de référence pour l'identification des souches de *Leishmania* au niveau spécifique et intra-spécifique. L'utilisation de cette méthode dans un cadre cladistique a également permis d'élaborer les premières hypothèses phylogénétiques (Rioux *et al.* 1990; Thomaz-Soccol *et al.* 1993), attestant en particulier la monophylie des sous-genres *Leishmania* et *Viannia*. L'analyse des isoenzymes a permis de confirmer la plupart des classifications préalables tout en y apportant quelques modifications, telles que la composition des complexes d'espèces ou la séparation des sections *Paraleishmania* et *Euleishmania* (Cupolillo *et al.* 2000), aboutissant à la classification actuellement reconnue, contenant plus de 50 espèces dont environ 20 sont connues pour infecter l'homme (Figure 19).

Durant les deux dernières décennies, l'analyse des séquences d'ADN de *Leishmania* a été de plus en plus utilisée à des fins d'identification et d'exploration phylogénétique. En comparaison des techniques enzymatiques, l'utilisation directe de l'ADN présente l'avantage de pouvoir être réalisée rapidement et sur de petites quantités de matériel inerte, sans culture préalable du parasite, tout en apportant une grande quantité d'informations (chaque nucléotide pouvant être considéré comme un caractère). Différentes méthodes d'identification ont été développées, se basant généralement sur l'amplification d'une région génomique par PCR. L'hybridation spécifique sur des sondes d'ADN marquées a été utilisée relativement tôt pour typer les souches de *Leishmania* (Barker & Butcher 1983; Degrave *et al.* 1994). L'analyse du

polymorphisme des fragments de restriction (PCR-RFLP; Marfurt *et al.* 2003; Rotureau 2006) rencontra ensuite beaucoup de succès et est utilisée encore aujourd'hui (Koarashi *et al.* 2016). De plus en plus, les méthodes comparatives basées sur le séquençage de marqueurs génétiques sont employées (Zelazny *et al.* 2005; Marco *et al.* 2006; de Almeida *et al.* 2011).

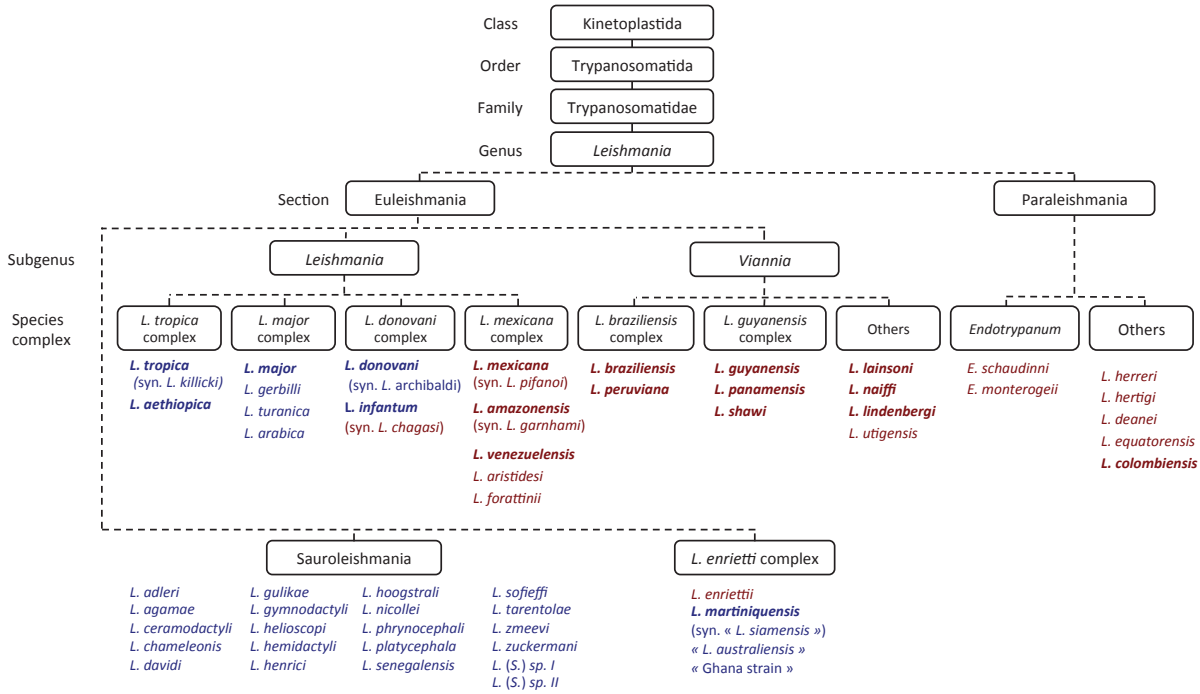


Figure 19: Classification des espèces du genre *Leishmania*. Les espèces présentes dans l'Ancien Monde sont indiquées en bleu, celles présentes dans le Nouveau Monde sont indiquées en rouge. Les espèces ayant été trouvées chez l'homme sont indiquées en gras (adapté de Akhoundi *et al.* 2016).

Toutefois, les régions génomiques ciblées ne permettent pas toujours de discriminer toutes les espèces de *Leishmania* telles qu'identifiées par l'analyse des isoenzymes. Ces données ont conduit à suggérer l'existence d'espèces synonymes telles que et *L. infantum* et *L. chagasi* (Kuhls *et al.* 2011), *L. donovani* et *L. archibaldi* (Jamjoom *et al.* 2004), *L. tropica* et *L. killicki* (Schönian *et al.* 2001), *L. mexicana* et *L. pifanoi* ou encore *L. amazonensis* et *L. garnhami*. La définition d'espèces chez ces organismes parasitaires est complexe et relève d'une certaine part de subjectivité (Tibayrenc 2006), mais les études génétiques appellent globalement à une simplification de la nomenclature des *Leishmania* (Schönian *et al.* 2010). Par ailleurs, l'utilisation de l'information génétique a permis beaucoup de progrès dans l'étude de la structure des populations de *Leishmania* (Bañuls *et al.* 2007; Rougeron *et al.* 2015) ainsi que dans la connaissance de l'histoire évolutive et biogéographique du groupe (Asato *et al.* 2009; Fraga *et al.* 2013; Harkins *et al.* 2016, Figure 20).

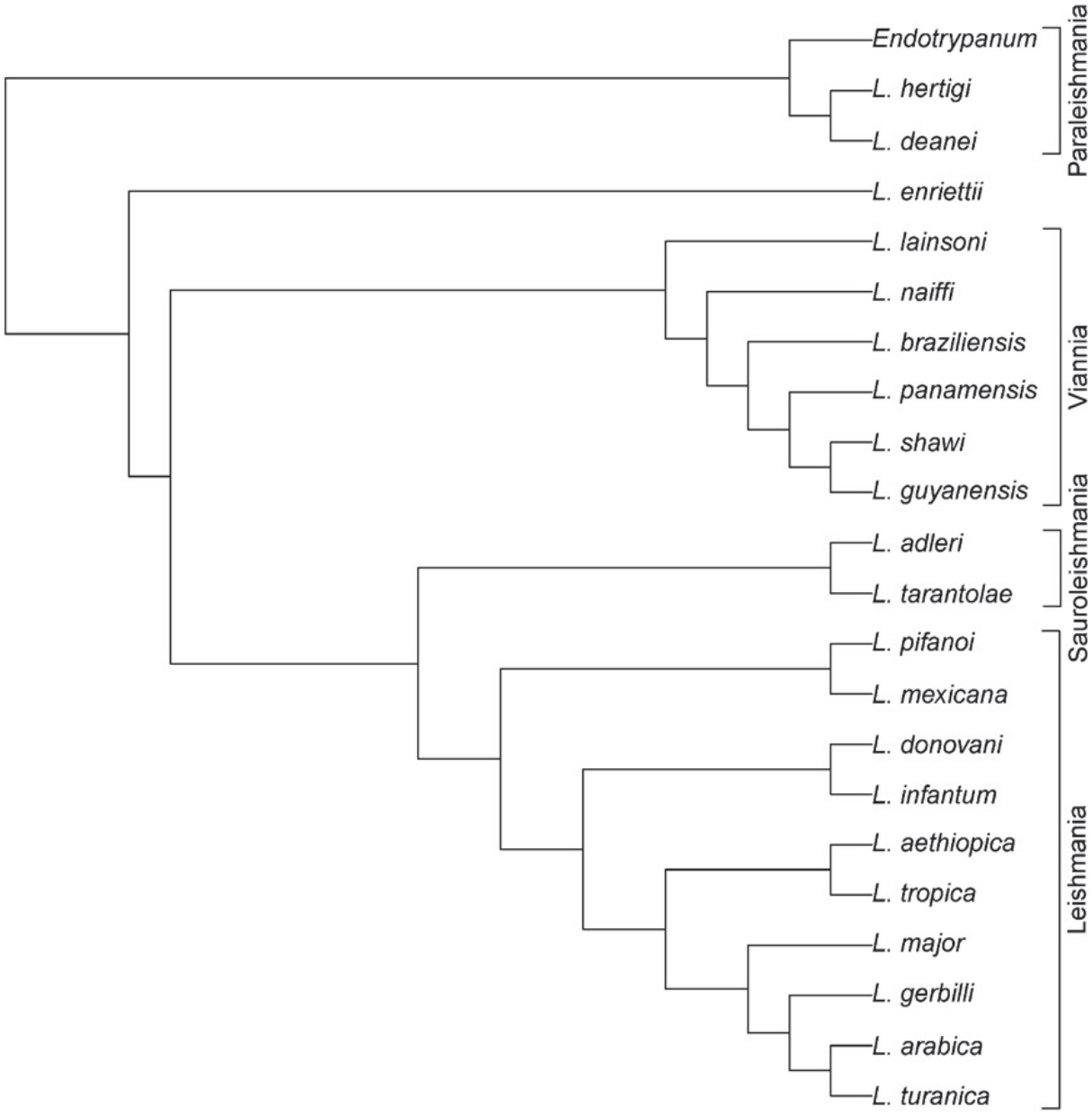


Figure 20: Arbre phylogénétique du genre *Leishmania* obtenu par méthode de maximum de vraisemblance sur 215 644 sites nucléotidiques variables sur le génome. Les supports de bootstrap sont de 100% pour tous les nœuds (Harkins *et al.* 2016).

5. LES PHLÉBOTOMES

A. BIOLOGIE

D'après Young & Duncan (1994), Killick-Kendrick (1999) et Maroli *et al.* (2013).

Les phlébotomes (Phlebotominae) sont de petits diptères hématophages de la famille des Psychodidae. Ils sont distribués sur les cinq continents, principalement dans les zones les plus chaudes (de 40° S à 50° N environ). Les formes adultes sont de petite taille (rarement >3 mm), couverts de soies assez denses, et d'une couleur pouvant varier de blanc à presque noir, bien que leur dénomination anglaise, « sand flies », fasse référence à une couleur sable pâle fréquemment rencontrée (Figure 21). Leurs ailes lancéolées sont maintenues dans une position typique en « V » au repos, et leurs pattes sont longues et fines.

Les individus des deux sexes se nourrissent de composés sucrés présents dans les tissus végétaux ou le nectar de fleurs, ainsi que de miellat de pucerons (Schlein & Muller 1995; Muller & Schlein 2004). Seules les femelles sont hématophages, le repas sanguin étant nécessaire au développement des œufs. Toutefois, quelques espèces, dîtes autogènes, peuvent accomplir leur première oviposition sans prise de repas sanguin. Il existe également des différences dans le nombre de repas sanguins devant être pris pendant chaque cycle gonotrophique selon les espèces.

Le cycle de développement des phlébotomes inclue plusieurs stades: œuf, larve (quatre stades larvaire), et adulte. La femelle dépose ses œufs dans des milieux humides et riches en matière organique tels que le sol de la forêt ou de bâtiments d'élevage, permettant le développement des larves détritivores sur le même site (Feliciangeli 2004). Le nombre d'œufs déposés peut varier d'une dizaine à une petite centaine, en fonction des espèces et des conditions (Ready 1979; Ferro *et al.* 1998; Kasap & Alten 2006). La durée complète d'un cycle varie typiquement de un à deux mois en laboratoire (Volf & Volfova 2011) et la durée de vie des adultes varie de une à trois semaines (Ferro *et al.* 1998; Schlein & Jacobson 1999; Kasap & Alten 2006), mais de longues phases de diapause sont observées chez l'œuf et les derniers stades larvaires pour survivre aux baisses de température hivernales ou à d'autres conditions défavorables (Killick-Kendrick & Killick-Kendrick 1987; Lawyer & Young 1991). Les sites de repos des adultes sont généralement des micro-habitats frais, humides et ombragés: à l'intérieur de bâtiments ou dans des fissures de murs en milieu anthropisé, sur le sol, sous les feuilles ou sur les troncs d'arbres en milieux sylvatiques mais aussi dans des caves ou des terriers d'animaux.

Contrairement aux moustiques, le vol des phlébotomes est silencieux. Il est également lent, ce qui les rend sensibles au vent et à la pluie (Colacicco-Mayhugh *et al.* 2011), et limite leur capacité de dispersion à de courtes distances, de l'ordre de la centaine de mètres au kilomètre en fonction de l'espèce et du milieu (Morrison *et al.* 1993; Casanova *et al.* 2005; Orshan *et al.* 2016). Les phlébotomes sont principalement actifs de la tombée de la nuit à l'aube, mais ils peuvent également piquer de jour lorsqu'ils sont dérangés.

Une saisonnalité dans l'abondance des phlébotomes peut être observée. Dans les zones tempérées, elle est généralement associée aux fluctuations de température avec une augmentation de l'abondance pendant l'été, montrant des pics unimodaux ou bimodaux (Ostfeld *et al.* 2004; Rossi *et al.* 2008). Dans les zones tropicales, elle est liée à la pluviométrie, mais les pics d'abondance sont observés en saison sèche ou en saison des pluies selon les études (Shaw & Lainson 1972; Oliveira *et al.* 2008). L'absence d'un patron de saisonnalité systématique reflète probablement la dépendance des phlébotomes à un compromis entre humidité et pluviométrie ou entre température et sècheresse dont l'optimum peut varier selon l'espèce et correspondre à différentes périodes saisonnières selon la localisation géographique.

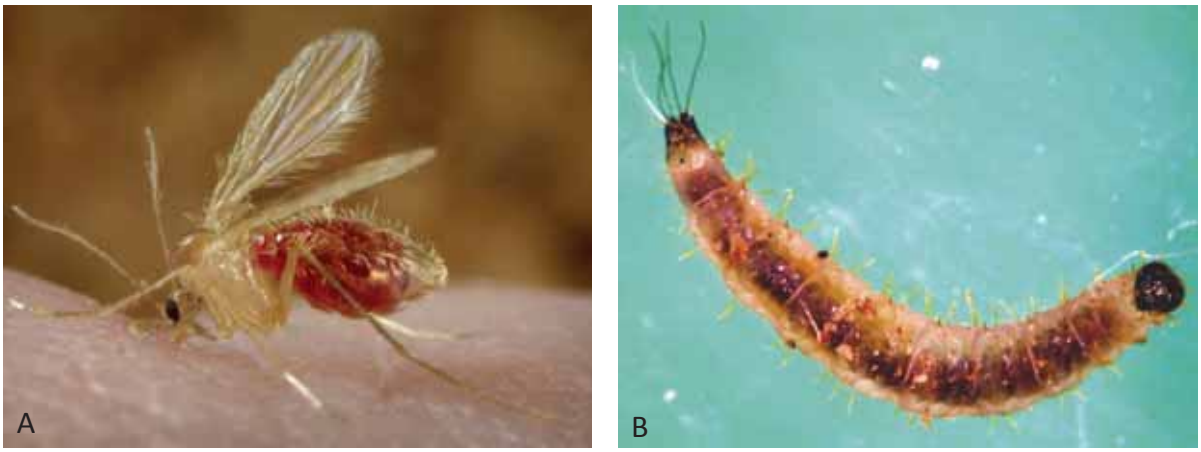


Figure 21: (A) femelle de phlébotome (*Phlebotomus papatasi*) prenant un repas sanguin (source: CDC); **(B)** quatrième stade larvaire de *P. perniciosus* (extrait de Maroli *et al.* 2013).

B. CAPACITÉS VECTORIELLES ET PRÉFÉRENCES TROPHIQUES

Notre capacité à prédire l'importance d'une espèce de phlébotome dans le cycle de transmission des *Leishmania* est limitée par une méconnaissance de l'histoire évolutive des deux groupes, l'apparition de coadaptations récentes, des variations intra et inter-populationnelles, et l'importance de facteurs écologiques locaux régissant les interactions entre vecteurs-hôtes (Ready 2013). Ainsi, bien que certains sous-genres de phlébotomes

(sensu Young & Duncan 1994) semblent être d'une importance épidémiologique particulière (*e.g. Nyssomyia*, *Lutzomyia*, groupe *Verrucarum* et *Psychodopygus* dans les néotropiques; Shaw *et al.* 1999); les recherches ne doivent pas se cantonner à ces groupes taxonomiques. L'incrimination d'une espèce vectrice d'importance médicale doit se faire sur la base de critères biologiques et épidémiologiques: observation d'individus infectés en conditions naturelles, démonstration expérimentale de la capacité à transmettre le parasite et mise en évidence de l'importance épidémiologique sur le terrain (Killick-Kendrick 1990; Ready 2013; Encadré 2). Actuellement, il existe une centaine d'espèces de phlébotomes prouvées ou suspectées vectrices de leishmaniose. Elles appartiennent principalement au genre *Phlebotomus* dans l'Ancien Monde et au genre *Lutzomyia* (*sensu* Young & Duncan 1994) dans le Nouveau Monde. Le genre *Sergentomyia* a toutefois été récemment incriminé dans l'Ancien Monde (Senghor *et al.* 2016).

Encadré 2: Critères d'incrimination d'une espèce de phlébotome vectrice d'importance médicale (adapté de Killick-Kendrick 1990)

1: Le vecteur est attiré et peut se nourrir sur un hôte réservoir (et sur l'homme dans le cas de la leishmaniose humaine)

2: A plus d'une occasion, les mêmes *Leishmania* sont isolées et identifiées chez des femelles sauvages et chez des hôtes mammifères

4: Une association spatio-temporelle existe entre le vecteur, les *Leishmania* et les hôtes réservoir.

5: Le parasite peut se multiplier expérimentalement chez le vecteur

6: Le vecteur peut transmettre expérimentalement le parasite à un hôte réservoir ou à un modèle de laboratoire équivalent.

Le rôle des différentes espèces de phlébotomes dans le cycle de transmission des *Leishmania* est intimement lié à leur propension à se nourrir sur les hôtes réservoirs ainsi que sur l'homme. Certaines espèces montrent des préférences trophiques très marquées, comme *Lu. vespertilionis* pour les chauve-souris (Christensen & Herrer 1980), *Lu. flaviscutellata* pour les rongeurs (Lainson & Shaw 1968), ou les espèces du sous-genre *Micropyga* pour les lézards (Tesh *et al.* 1971). La plupart sont toutefois considérées comme généralistes, pouvant se nourrir selon les opportunités sur une grande diversité d'oiseaux et de mammifères (Ready 2013). Ainsi, le caractère anthropophile de certaines espèces est probablement plus lié à une adaptation aux milieux domestiques et à d'autres caractéristiques écologiques qui favorisent leur mise en contact avec l'homme plutôt qu'à une réelle préférence trophique (de Oliveira *et al.* 2008; Jaouadi *et al.* 2013).

C. CLASSIFICATION ET IDENTIFICATION

D'après Bates et al. (2015) et Akhoundi et al. (2016).

Les phlébotomes sont des diptères nématocères de la famille des Psychodidae (sous-famille: Phlebotominae) comprenant près de 1000 espèces distribuées sur tous les continents excepté l'Antarctique, et dont plus de la moitié se trouve en Amérique (Bates *et al.* 2015). Depuis la première description reconnaissable d'un phlébotome par Bonanni en 1691 (considéré par lui comme une espèce de moustique), et la description de *Phlebotomus papatasi* par Scopoli en 1786 comme la première espèce de Psychodidae, la classification des phlébotomes s'est faite historiquement selon une approche phénétique, regroupant les taxons sur la base de similarités morphologiques. Jusqu'au début du 20ème siècle, des caractères morphologiques externes tels que les pièces génitales des mâles ou les nervures alaires étaient utilisés. Par la suite, l'analyse de la structure d'organes internes tels que les spermathèques, le cibarium ou le pharynx fut également employée (Figure 22).

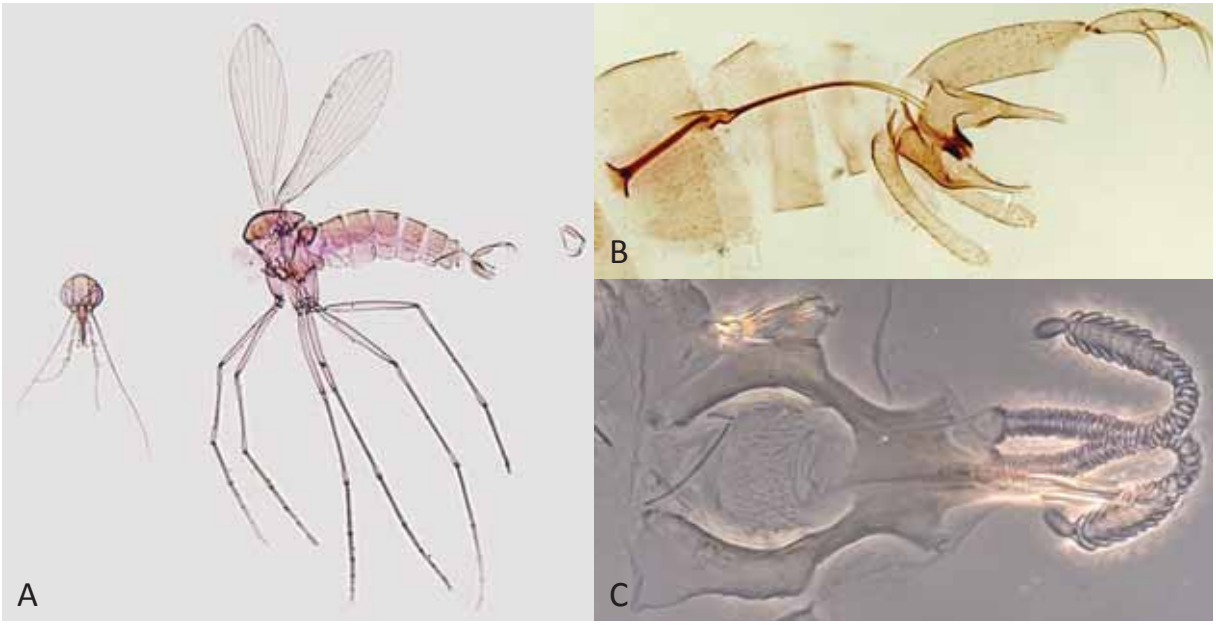


Figure 22: (A) Individu mâle entier de *Lutzomyia cayennensis* monté sur lame; **(B)** Pièces génitales mâles de *Lu. umbratilis*; **(C)** Pièces génitales femelles de *Lu. s. maripaensis* (clichés: Jean-Charles Gantier).

Beaucoup de systèmes de classification se sont succédé incluant ceux de Newstead, Theodor, Abonnenc et Lewis. En 1977, Lewis *et al.* proposent un système de classification repris ensuite avec quelques modifications par Young & Duncan (1994), et qui est encore largement utilisé aujourd'hui. Dans ce système, les phlébotomes sont répartis en six genres, dont trois dans l'ancien monde (*Phlebotomus*, *Sergentomyia* et *Chinius*), et trois dans le nouveau monde (*Lutzomyia*, *Brumptomyia* et *Warileya*).

Le genre *Phlebotomus* comprend une centaine d'espèces, dont toutes celles incriminées dans la transmission de la leishmaniose humaine dans l'ancien monde. La plupart sont trouvées en milieux ouverts et semi-arides plutôt qu'en forêt. Le genre *Sergentomyia* inclut plus de 300 espèces, et domine dans les régions tropicales où le genre *Phlebotomus* est plus rare. Les phlébotomes du genre *Sergentomyia* se nourrissent principalement sur des vertébrés à sang froid et ne sont pas considérés comme vecteurs de leishmaniose humaine, même si plusieurs études remettent leurs préférences trophiques et ce paradigme en question (*e.g.* Jaouadi *et al.* 2013, Senghor et al. 2016). Le genre *Chinius* ne comprend actuellement que quatre espèces cavernicoles d'Asie du Sud-Est.

En Amérique, la grande majorité des espèces de phlébotomes appartiennent au genre *Lutzomyia* (près de 500 espèces). En particulier, ce genre contient toutes les espèces vectrices de leishmaniose. Dans cette région, les phlébotomes sont essentiellement associés au milieu tropical et sylvatique, bien que certaines espèces comme *Lu. longipalpis* semblent bien s'adapter au milieu péridomestique (Rangel & Vilela 2008). Le genre *Brumptomyia* comprend une trentaine d'espèces souvent associées aux terriers de tatous, tandis que le genre *Warileya* ne contient qu'une petite dizaine d'espèces dont l'écologie est mal connue.

La connaissance actuelle de l'histoire évolutive des phlébotomes est mince, et repose essentiellement sur une analyse cladistique de caractères morphologiques (Galati 1995). Les résultats de Galati suggèrent la paraphylie des genres *Lutzomyia* et *Phlebotomus*. L'auteur propose un nouveau système de classification basé sur des critères phylogénétiques (qui sera réactualisé jusqu'à récemment; Galati 2014), dans lequel les Phlebotominae sont divisés en deux tribus et six sous-tribus, et beaucoup de sous-genres de *Lutzomyia sensu* Young & Duncan (1994) sont élevés au rang générique (Figure 23). Quelques travaux basés sur l'analyse de séquences d'ADN ont été réalisés pour des groupes restreints (Depaquit 2014), mais aucune étude moléculaire à grande échelle n'a été conduite à ce jour pour conforter les résultats précédemment obtenus par l'analyse manuelle (*i.e.* sans assistance informatique) de caractères morphologiques. La classification de Galati est donc, aujourd'hui encore, la seule hypothèse phylogénétique substantielle existante. Malgré cela, la classification de Young & Duncan (1994) reste largement la plus employée pour des raisons pratiques (nombre restreint de genre et cohérence historique).

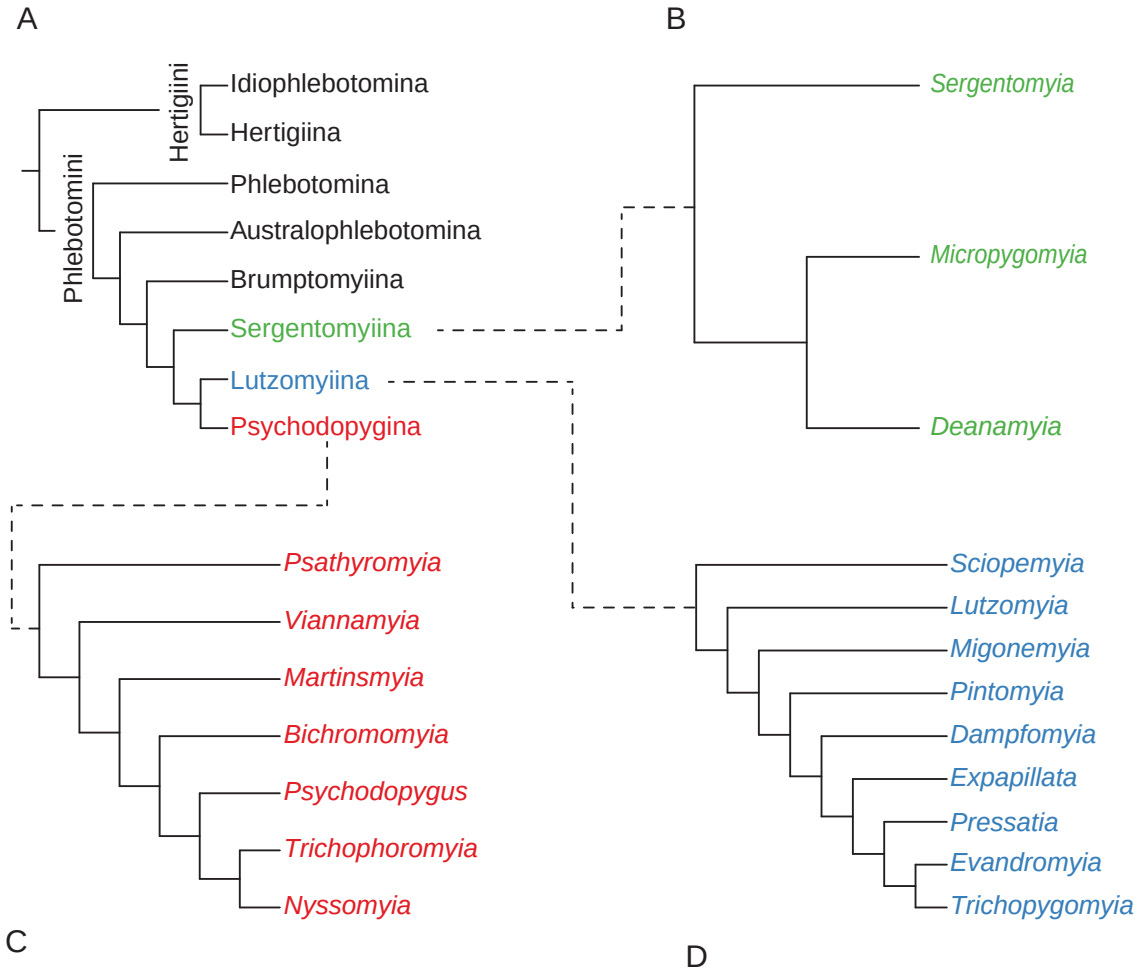


Figure 23: Version simplifiée de la classification des phlébotomes proposée par Galati (1995). **(A)** Relations phylogénétiques entre les différentes tribus et sous-tribus de Phlebotominae; **(B,C,D)** relations phylogénétiques entre les différents genres compris dans les sous-tribus Sergentomyiina, Psychodopygina et Lutzomyiina.

Les méthodes traditionnelles d'identifications des phlébotomes basées sur des clés de détermination morphologique (Young & Duncan 1994; El-Hossary 2006) nécessitent des manipulations délicates et chronophages (dissection, traitement chimique, montage sur lame), ainsi qu'une expertise taxonomique solide. Différentes méthodes moléculaires ont été développées pour s'affranchir de cette tâche. Les deux techniques principalement employées sont la PCR-RFLP (Aransay *et al.* 1999; Terayama *et al.* 2008; Minter *et al.* 2013) et le barcoding d'ADN (Krüger *et al.* 2011; Contreras Gutiérrez *et al.* 2014; Romero-Ricardo *et al.* 2016). Le barcoding d'ADN consiste à séquencer un fragment d'ADN standard amplifié par PCR pour réaliser l'assignation taxonomique par comparaison à des séquences de référence. Ces outils rencontrent un succès grandissant, car ils permettent l'identification de spécimens sans expertise taxonomique. Réciproquement, des inquiétudes peuvent être émises quand au devenir de l'alpha-taxonomie, qui est à la base de toutes les autres études, et dont les spécialistes se font de plus en plus rares.

6. ÉCO-ÉPIDEMIOLOGIE DES LEISHMANIOSES

A. SYSTÈMES ÉPIDÉMIOLOGIQUES

D'après Shaw (2002), Lainson & Shaw (2005), Rotureau (2006), Maroli *et al.* (2013) et Ready (2013).

Dans l'Ancien Monde, trois espèces de *Leishmania* sont caractérisées par un cycle de transmission zoonotique. *L. infantum*, agent de la LV et de LC, est largement répandue dans les pays du bassin méditerranéen et de la péninsule arabique ainsi qu'en Asie centrale et en Chine. Le parasite est transmis par différentes espèces de phlébotomes appartenant majoritairement au genre *Phlebotomus* telles que *P. ariasi* ou *P. perniciosus* et possiblement au *Sergentomyia dubia* et *S. schewtzi* (Senghor *et al.* 2016). Le chien représente le réservoir le plus important en milieu domestique auquel certaines espèces vectrices s'adaptent bien (Rossi *et al.* 2008; Guernaoui & Boumezzough 2009), mais le parasite circule également chez différents carnivores sauvages dont le loup et le renard (Mohebali *et al.* 2005; Sobrino *et al.* 2008). La LC classique est causée par *L. major,* typique des milieux arides et de savanes de l'Afrique du Nord jusqu'en Asie du Sud, où différentes espèces de rongeurs sauvages jouent le rôle de réservoirs. Les vecteurs incriminés dans la transmission de *L. major* appartiennent au sous-genre *Phlebotomus (Phlebotomus)* (*P. papatasi* et *P. duboscqi* en particulier). Un autre agent de LC, *L. aethiopica* est restreint aux zones d'altitude de l'Ethiopie et du Kenya. Il circule chez les hyracoïdes (e.g. Daman du Cap), et est transmis par *P.* (*La.*) *longipes* et *P.* (*La.*) *pedifer.*

Deux autres espèces de *Leishmania* de l'Ancien Monde sont caractérisées par des cycles de transmission strictement ou essentiellement anthroponotiques. *L. donovani,* agent de la LV, est concentré en Asie dans le nord de l'Inde, au Bangladesh et au Népal où le seul vecteur incriminé est *P.* (*Euphlebotomus*) *argentipes*. Elle est également présent en Chine, dans une moindre mesure (Lun *et al.* 2015), et probablement aussi dans certains pays de la péninsule arabique où peu de données sont disponibles (Postigo 2010). *L. donovani* circule également en Afrique de l'Est (Kenya, Ethiopie, Somalie, Soudan et Ouganda) où il est transmis par *P.* (*La.*) *orientalis* et *P.* (*Symphlebotomus*) *martini. L. tropica* est responsable d'une LC principalement anthroponotique rencontrée principalement dans les régions semi-arides du sud de la Grèce au nord-ouest de l'Inde, et est typiquement transmise par *P.* (*Paraphlebotomus*) *sergenti* en milieu urbain. Elle est également présente dans des foyers plus épars en Afrique et au Moyen-Orient où elle est transmise par d'autres espèces de

phlébotomes, en milieu rural, et a occasionnellement été observée chez des mammifères sauvages (Sang *et al.* 1992; Svobodova *et al.* 2006; Jaouadi *et al.* 2013).

Dans le nouveau monde, les leishmanioses sont distribuées de l'Argentine jusqu'au Texas et sont toutes zoonotiques. La LV est causée par *L. chagasi* (*syn. L. infantum*) qui aurait été importée récemment par des chiens infectés accompagnant les colons européens sur le continent Américain (Kuhls *et al.* 2011). Le parasite a trouvé sur place une espèce vectrice locale: *Lutzomyia longipalpis.* On sait aujourd'hui que *Lu. longipalpis* est un complexe comprenant plusieurs espèces dont le rôle épidémiologique varie (Arrivillaga & Feliciangeli 2001), et que *L. chagasi* est transmis par d'autres espèces de phlébotomes comme *Lu. evansi* en Colombie (Travi *et al.* 1996). La plupart des cas de LV sont concentrés au Brésil, mais le parasite se retrouve largement sur le continent en milieu tropical sec. Le cycle de transmission a lieu dans les zones péridomestiques et domestiques auxquelles *Lu. longipalpis* semble bien adaptée, et où le chien joue le rôle de réservoir. Comme dans l'Ancien Monde, d'autres carnivores sauvages peuvent héberger *L. chagasi* en Amérique. Elle a par exemple été retrouvée chez le renard des savanes au Brésil (Lainson *et al.* 1990) et l'opossum commun en Colombie (Travi *et al.* 1994). Des cycles strictement sauvages existent probablement en milieu sylvatique où la présence de *Lu. longipalpis* est également possible (Lainson *et al.* 1990).

La LC est causée par un grand nombre d'espèces de *Leishmania* différentes sur le continent américain, dont la plupart sont caractérisées par un cycle de transmission sylvatique impliquant différents mammifères sauvages. Ainsi, la maladie est classiquement associée aux activités forestières comme en atteste ses différents noms vernaculaires sur le continent: « ulcère des chicleros » (travailleurs récoltant le latex d'arbres forestiers) au Mexique, « plaie des guérilléros » au Venezuela et en Colombie, « pian-bois » en créole guyanais. Toutefois, la LC s'est étendue aux zones péridomestiques depuis les années 1960, à la fois du fait de l'incursion de l'habitat humain en milieu forestier dans certaines régions et de l'adaptation de certaines espèces d'hôtes et de vecteurs aux milieux déforestés et agricoles (Campbell-Lendrum *et al.* 2001; Travi *et al.* 2002; Quaresma *et al.* 2011). Les réservoirs sauvages de LC souffrent rarement de lésions. Ce n'est pas le cas pour certains mammifères domestiques tels que le chien ou le cheval qui peuvent occasionnellement présenter d'importants ulcères cutanés suite à une infection par *L. braziliensis.* Ceci a tendance à les faire considérer comme des hôtes accidentels, qui tout comme l'homme, ne permettent pas à eux seuls de maintenir la circulation du parasite. Les particularités écologiques des espèces de *Leishmania* responsables de LC dans le Nouveau Monde sont résumées dans le Tableau 3.

Tableau 3: Caractéristiques éco-épidémiologiques des différentes espèces de Leishmania responsables des leishmanioses cutanées sur le continent américain (d'après (Shaw 2002; Rotureau 2006a; Lainson & Shaw 2010; Ready 2013).

Espèce	**Distribution géographique**	**Hôtes non-humains**	**Vecteurs**	**Ecologie**
L.* (*Leishmania*) *mexicana (complexe *L. mexicana)*	Sud des Etats-Unis et Amérique centrale	Rongeurs forestiers (Cricetidae)	Prouvés: *Lutzomyia* (*Nyssomyia*) *olmeca olmeca* Suspectés: *Lu.* (*Ny.*) *olmeca bicolor* *Lu.* (*Dampfomyia*) *anthophora* *Lu.* (*Lutzomyia*) *diabolica*	Peu d'études décrivent l'écologie de ce parasite. Le cycle de transmission est sylvatique, toutefois, les vecteurs et les réservoirs semblent bien s'adapter aux habitats perturbés.
L.* (*L.*) *amazonensis (complexe *L. mexicana)*	Région amazonienne	Rongeurs du genre *Proechimys* (hôte principal), mais également d'autres rongeurs (Cricetidae, Dasyproctydae) et certains marsupiaux.	Prouvés: *Lutzomyia* (*Ny.*) *flaviscutellata* Suspectés: *Lu.* (*Ny.*) *olmeca nociva* *Lu.* (*Ny.*) *olmeca reducta* *Lu.* (*Lu.*) *longipalpis*	Les hôtes et vecteurs principaux sont présents dans différents habitats forestiers, incluant des habitats exploités ainsi que dans des plantations. *Lu. flaviscutellata* est présent au niveau du sol et semble afficher une préférence trophique nette pour les rongeurs, bien qu'il puisse se nourrir occasionnellement sur une grande diversité de vertébrés incluant l'homme.
L.* (*L.*) *venezuelensis (complexe *L. mexicana)*	Nord du Venezuela	*L. venezuelensis* a été retrouvée chez certains mammifères domestiques (chat, cheval), mais son réservoir sauvage reste indéterminé.	Prouvés: Aucun Suspectés: *Lu.* (*Ny.*) *olmeca bicolor*	Peu d'informations sont disponibles concernant *L. venezuelensis*, mais sa transmission semble avoir lieu dans des habitats très variés, suggérant que ses réservoirs sauvages soient de petits mammifères pouvant s'adapter aux milieux péridomestiques.

Tableau 3 (suite) : Caractéristiques éco-épidémiologiques des différentes espèces de Leishmania responsables des leishmanioses cutanées sur le continent américain (d'après (Shaw 2002; Rotureau 2006a; Lainson & Shaw 2010; Ready 2013).

Espèce	**Distribution géographique**	**Hôtes non-humains**	**Vecteurs**	**Ecologie**
L.* (*Viannia*) *braziliensis (complexe *L. braziliensis)*	La distribution géographique de *L. braziliensis* est mal définie du fait des difficultés dans l'identification de ce parasite qui présente un haut degré de polymorphisme, mais des *Leishmania* appartenant au complexe *L. braziliensis* ont été décrites dans la plupart des pays d'Amérique latine.	Diverses espèces de rongeurs et marsupiaux, ainsi que certains mammifères domestiques présentant des manifestations cliniques (chien, cheval, mule).	Prouvés: *Lutzomyia* (*Psychodopygus*) *wellcomei* *Lu.* (*Ny.*) *neivai* *Lu.* (*Ny.*) *whitmani* *Lutzomyia* (*Pifanomyia*) *ovallesi* Suspectés: *Lu.* (*Ny.*) *intermedia* *Lu.* (*Migonei*) *migonei* *Lu.* (*Ps.*) *carrerai* *Lu.* (*Pf.*) *towsendis* *Lu.* (*Helcocyrtomyia*) *pescei* *Lu.* (*Ps.*) *panamensis*	L'écologie de *L. braziliensis* apparait complexe et mosaïque, du fait du polymorphisme de ce parasite, de la diversité d'hôtes et de vecteurs impliqués dans sa transmission. Un patron qui semble ressortir est que *L. braziliensis* est transmis dans la forêt amazonienne primaire par des phlébotomes du sous-genre *Psychodopygus*, tandis que dans d'autres régions, il est transmis dans une grande variété d'habitats incluant des habitats forestiers perturbés et péridomestiques par des espèces de phlébotomes adaptés à ces milieux (du sous-genre *Nyssomyia* en particulier), différentes espèces de petits mammifères réservoirs permettant de faire le lien entre ces habitats. Les réservoirs sauvages principaux n'ont toujours pas été identifiés clairement, et le degré d'implication de la faune domestique dans sa transmission n'est pas bien défini.
L.* (*V.*) *peruviana (complexe *L. braziliensis)*	Région andine du Pérou de l'Argentine et de l'Equateur.	Les réservoirs de *L. peruviana* sont mal identifiés, mais il a été détecté chez une espèce de rongeur (*Phyllotis andinum*) et chez l'opossum commun ainsi que chez le chien.	Prouvés: Aucun Suspectés: *Lu.* (*He.*) *peruensis* *Lu.* (*Pf.*) *verrucarum*	*L. peruviana* circule dans les zones sèches et montagneuses, en milieu sylvatique et péridomestique. Le rôle du chien dans sa transmission n'est pas clair.

Tableau 3 (suite) : Caractéristiques éco-épidémiologiques des différentes espèces de Leishmania responsables des leishmanioses cutanées sur le continent américain (d'après (Shaw 2002; Rotureau 2006a; Lainson & Shaw 2010; Ready 2013).

Espèce	Distribution géographique	Hôtes non-humains	Vecteurs	Ecologie
L.* (*V.*) *guyanensis (complexe *L. guyanensis*)	Région amazonienne	Les hôtes principaux sont des xénarthres arboricoles (paresseux à deux doigts et fourmilier à collier), mais le parasite à également été isolé chez certains rongeurs du genre *Proechimys*, chez l'opossum commun (*Didelphis marsupialis*) et chez le kinkajou (*Potos flavus*).	Prouvés: *Lutzomyia* (*Ny.*) *umbratilis* Suspectés: *Lu.* (*Ny.*) *anduzei* *L.* (*Ny.*) *whittmani* *Lu.* (*Ny.*) *shawi*	Le cycle de transmission de *L. guyanensis* est typiquement sylvatique. Le vecteur *Lu. umbratilis* est actif de nuit dans la canopée où les hôtes arboricoles sont également présents. Durant le jour, on peut le trouver à la base des troncs d'arbres, et il se nourrit fréquemment sur l'homme lorsqu'il est dérangé.
L.* (*V.*) *panamensis (complexe *L. guyanensis*)	Panama essentiellement, mais le parasite a été détecté du Honduras à l'Equateur et au Venezuela.	Le paresseux d'Hoffmann (*Choloepus hoffmanni*) est considéré comme l'hôte principal, mais le parasite a occasionnellement été détecté chez d'autres mammifères sylvatiques (procyonidés, primates et rongeurs).	Prouvés: Aucun Suspectés: *Lu.* (*Ny.*) *trapidei* *Lu.* (*Ny.*) *ylephiletor* *Lu.* (*Ny.*) *edentulata* *Lu.* (*Lu.*) *gomezi*	*L. panamensis* est très proche de *L. guyanensis* (si bien qu'il a été suggéré qu'il s'agisse de la même espèce), et présente des caractéristiques écologiques très semblables avec un cycle exclusivement sylvatique, des hôtes réservoirs et des espèces vectrices arboricoles.
L.* (*V.*) *shawi (complexe *L. guyanensis*)	Nord du Brésil	*L. shawi* a été isolé chez des primates non-humains (*Cebus apella, Chiropotes satanas*), chez des xenarthres (*Choloepus didactylus* et *Bradypus tridactylus*) et chez le coati (*Nasua nasua*).	Prouvés: Aucun Suspectés: *Lu.* (*Ny.*) *whitmani*	Les connaissances écologiques de *L. shawi* reposent essentiellement sur une seule étude. Le parasite circule probablement chez d'autres mammifères arboricoles. Il n'a jusqu'ici été trouvé que chez une espèce de phlébotome identifiée comme *Lu.* (*Ny.*) *whitmani*, mais des données génétiques et morphométriques suggèrent qu'il puisse s'agir d'une espèce cryptique, caractérisée par un habitat strictement sylvatique et une faible anthropophilie, en opposition avec le caractère péridomestique et antrhopophile de cette espèce dans d'autres régions du Brésil.

Tableau 3 (suite): Caractéristiques éco-épidémiologiques des différentes espèces de Leishmania responsables des leishmanioses cutanées sur le continent américain (d'après (Shaw 2002; Rotureau 2006a; Lainson & Shaw 2010; Ready 2013)**.**

Espèce	**Distribution géographique**	**Hôtes non-humains**	**Vecteurs**	**Ecologie**
L.* (*V.*) *lainsoni	Région amazonienne	Le seul hôte non-humain connu de *L. lainsoni* est le paca (*Cuniculus paca*).	Prouvés: Aucun Suspectés: *Lu.* (*Trichophoromyia*) *ubiquitalis* *Lu.* (*Pf.*) *nuneztovari* *Lu.* (*Th.*) *ininii*	*L. lainsoni* circule en milieu sylvatique. Les cas d'infection chez l'homme sont rares, probablement du fait d'une faible anthropophilie du vecteur.
L.* (*V.*) *naiffi	Nord du Brésil, Guyane Française, Panama, Pérou, Equateur, Martinique	Le seul hôte non-humain connu de *L. naiffi* est le tatou à neuf bandes (*Dasypus novemcinctus*).	Prouvés: Aucun Suspectés: *Lu.* (*Ps.*) *ayrozai* (principalement) *Lu.* (*Ps.*) *paraensis* *Lu.* (*Ps.*) *s. squamiventris* *Lu.* (*Ps.*) *s. maripaensis*	*L. naiffi* circule en milieu sylvatique. Les vecteurs ne semblent pas être associés aux terriers de tatou, le cycle doit donc avoir lieu à l'extérieur. *Lu.* (*Ps.*) *ayrozai* semble peu anthropophile, expliquant peut être le faible nombre de cas humains. D'autres espèces vectrice suspectées (*Lu.* (*Ps.*) *paraensis*, *Lu.* (*Ps.*) *s. squamiventris*) semblent toutefois se nourrir plus volontiers sur l'homme.
L.* (*V.*) *lindenbergi	Etat de Para au Brésil	*L. lindenbergi* n'a été trouvé que chez l'homme jusqu'à présent.	Prouvés: Aucun Suspectés: *Lu.* (*Ny.*) *antunesi*	Très peu d'informations sont disponibles concernant *L. lindenbergi*. Le parasite n'a jusqu'ici été retrouvé que chez des soldats s'entrainant dans des zones de forêt dégradée autour de la ville de Belem.
L.* (Paraleishmania) *colombiensis	Colombie, Venezuela, Panama	Le seul hôte sauvage connu est le paresseux d'Hoffman (*Choloepus hoffmani*).	Prouvés: Aucun Suspectés: *Lu.* (*He.*) *hartmanni* *Lu.* (*Ps.*) *panamensis* *Lu.* (*Lu.*) *gomezi*	La distribution géographique large du parasite indique qu'il circule probablement chez d'autres mammifères sauvages, mais son cycle semble principalement arboricole. Il n'a été retrouvé que très rarement chez l'homme.

B. LES LEISHMANIOSES EN GUYANE

D'après (Dedet 1990; Rotureau 2005a)

La Guyane est un département-région d'outre mer situé en Amérique du Sud, au nord du Brésil, et tourné vers la zone Caraïbe. Avec ses 83 846 km^2 et 262 527 habitants (estimations INSEE 2016) c'est le plus grand département Français, mais également le moins peuplé (3.13 habitant/km^2). Il faut dire que le territoire est recouvert à 98% de forêt tropicale humide (World Resources Institute 2014), et que la population est concentrée sur la bande littorale. La Guyane bénéficie encore d'un statut de conservation favorable ce qui en fait un lieu privilégié pour l'étude de la biodiversité. Celle-ci est toutefois menacée par endroit du fait de l'importante pression démographique que connait la région (l'accroissement de la population guyanaise est de loin le plus élevé de France avec un taux annuel moyen supérieur à 2%), et de l'activité minière illégale qui a lieu dans l'intérieur des terres (Hammond *et al.* 2007).

La leishmaniose cutanée est une affection connue depuis longtemps des habitants de Guyane (Dedet *et al.* 1989b). Le premier cas diagnostiqué fut décrit en 1943 par Floch (Floch 1943), et le parasite nommé *Leishmania guyanensis* (Floch 1954). Depuis, quatre autres espèces de *Leishmania* ont été identifiées en Guyane: *L. amazonensis* (Dedet *et al.* 1984, 1985), *L. braziliensis* (Garin *et al.* 1989), *L. naiffi* (Darie *et al.* 1995) et *L. lainsoni* (Basset *et al.* 2001). Actuellement, plus d'une centaine de cas de leishmaniose cutanée sont répertoriés dans les hôpitaux de Guyane chaque année (Simon *et al.* 2017), mais il existe probablement de nombreux cas non diagnostiqués ou contractés lors d'un voyage temporaire qui ne sont pas pris en compte dans ses chiffres. Si *L. guyanensis* est toujours l'agent de la leishmaniose humaine majeur en Guyane (96,7% des cas en 1989, 83,7% en 2013; Desjeux & Dedet 1989; Simon *et al.* 2017), *L. braziliensis* semble progresser de façon inquiétante depuis plusieurs années (Dedet *et al.* 1994; Martin-Blondel *et al.* 2015; Simon *et al.* 2017).

Suite à une augmentation du nombre de cas à la fin des années 1970, un programme de recherche regroupant l'Institut Pasteur, l'ORSTOM (IRD aujourd'hui) et de l'hôpital de Cayenne fut lancé. Pendant près de 10 ans, la Guyane est devenue le siège de recherches transversales pionnières sur les leishmanioses, regroupant des approches éco-épidémiologiques et médicales. Les résultats de ces études, synthétisés par (Dedet 1990) ont permis de considérables avancées dans la connaissance des réservoirs et des vecteurs de leishmanioses (voir aussi Rotureau 2006), ainsi que dans leur prise en charge clinique et thérapeutique. Par ailleurs, la présence de militaires en forêt guyanaise et le travail des

services de médecine des armées ont permis de recueillir de précieuses données épidémiologiques (Banzet 2000; Lightburn *et al.* 2001).

Les travaux de Dedet et coll. ont constitué un premier bloc d'études écologiques des leishmanioses en Guyane (Tableau 4). En particulier, ils ont permis l'établissement de connaissances fondamentales sur le vecteur principal de la région: *Lu. umbratilis*. Ce n'est que près de 20 ans après que de nouvelles études écologiques seront menées, dans le cadre de la thèse de Rotureau (2005b), motivée par le besoin d'actualiser les connaissances suite aux changements environnementaux qu'a connu le territoire depuis les années 1980. En 2007, Fouque *et al.* identifient de nouvelles associations vecteur-parasite, une étude qui constitue la dernière du genre en Guyane.

Le présent travail de thèse représente donc, au delà des questions plus générales auxquelles il vise à répondre, la première source de données écologiques sur les leishmanioses en Guyane depuis ces dix dernières années.

Tableau 4: Synthèse (non-exhaustive) des études éco-épidémiologiques des leishmanioses réalisées en Guyane depuis le début des années 80.

Auteurs	**Année**	**Titre**	**Résumé**
Le Pont *et al.*	1980	Preliminary observations on the sylvatic cycle of leishmaniasis in French Guiana	Plusieurs spécimens de *Lutzomyia umbratilis* sont trouvés infectés par des *Leishmania*. Cette espèce est fortement suspectée d'être le vecteur de *L. guyanensis* en Guyane (comme cela a déjà été montré au Brésil). L'existence d'un cycle sylvatique similaire à celui de *L. panamensis* est probable. Deux rongeurs du genre *Proechimys* ont été trouvés infectés par des *Leishmania* non identifiées.
Le Pont & Pajot	1980	Leishmaniose en Guyane Française. 1. Etude de l'écologie et du taux d'infection naturelle du vecteur *Lutzomyia* (*Nyssomyia*) *umbratilis* Ward et Fraiha, 1977 en saison sèche. Considérations épidémiologiques	Les auteurs étudient l'écologie de *Lu. umbratilis* par des piégeages sur appât humain dans le deuxième semestre de l'année 1979 sur la piste de Saint-Elie. Ils montrent que: - L'abondance de *Lu. umbratilis* augmente du sol vers la canopée (<35% des phlébotomes anthropophiles au sol, >75% en Canopée). Sa répartition verticale se décale vers le bas en saison des pluies avec un pic d'activité au sol en Novembre. - Certains arbres sont plus propices que d'autres à héberger le vecteur. - Son activité est principalement crépusculaire ou post-crépusculaire, mais le vecteur est sensible à des intrusions diurnes. - Il est anthropophile mais semble opportuniste. Il se gorge très facilement sur le paresseux. - Le taux d'infection par des *Leishmania* varie grandement (20% en novembre, autour de 1.3% en décembre). - Le nombre de piqûres infestantes par homme-heure varie de 0.2 en septembre à 3.7 fin novembre
Gentile *et al.*	1981	Dermal leishmaniasis in French Guiana: the sloth (*Choloepus didactylus*) as a reservoir host	Les *Leishmania* sont recherchées chez 75 mammifères appartenant à 15 espèces. *L. guyanensis* est isolée chez 1/7 kinkajou (*Potos flavus*), 7/15 paresseux à deux doigts (*Choloepus didactylus*) et 2/19 rongeur du genre *Proechimys*.
Le Pont & Pajot	1981	La leishmaniose en Guyane française: 2. Modalités de la transmission dans un village forestier : Cacao	Les auteurs mènent une étude épidémiologique dans le village de Cacao où une flambée de cas a eu lieu dans les année 1979-80. Ils montrent que les cas sont associés à des lambeaux de forêt primaire qui persistent proches des habitations et concluent que la circulation du parasite peut se maintenir même dans une zone forestière fortement dégradée.

Tableau 4 (suite) : Synthèse (non-exhaustive) des études éco-épidémiologiques des leishmanioses réalisées en Guyane depuis le début des années 80.

Auteurs	Année	Titre	Résumé
Le Pont	1982	La leishmaniose en Guyane française: 2. Fluctuations saisonnières d'abondance et du taux d'infection naturelle de *Lutzomyia* (*Nyssomyia*) *umbratilis* Ward et Fraiha, 1977	Suivi de la population de *Lu. umbratilis* sur la piste de Saint-Elie pendant un an (extension de l'étude de 1980). L'abondance des phlébotomes est maximum en débit et fin de saison des pluies. Le risque infectieux cours de septembre à mars mais est maximal à la reprise des pluies. Sur un an, les deux tiers des phlébotomes capturés au sol sur homme appartiennent au genre *Psychodopygus*
Chippaux & Pajot	1983	La leishmaniose en Guyane française. IV: Note préliminaire sur les phlébotomes des terriers	*Lu. trichopyga* représente la majorité des phlébotomes piégés dans des terriers dans des zones forestières proches de Cayenne, mais d'autres espèces sont retrouvées, y compris *Lu. umbratilis*.
Dedet *et al.*	1984	Natural hosts of *Leishmania mexicana amazonensis* Lainson and Shaw, 1972 (Kinetoplastida: Trypanosomatidae) in French Guiana	*L. amazonensis* est recherché chez *Lu. flaviscutellata* ainsi que chez différentes espèces de mammifères. Le parasite est retrouvé chez 1/254 phlébotomes collectés et chez 2/22 rongeurs du genre *Proechimys*.
Geoffroy *et al.*	1986	Note sur les relations des vecteurs de leishmaniose avec les essences forestières en Guyane française	Des phlébotomes sont capturés sur les troncs de différentes essences d'arbres. Les phlébotomes retrouvés sont essentiellement *Lu. umbratilis* et *Lu. rorotaensis*. Des phlébotomes sont capturés sur toutes les essences, il y a des différences mais elles sont difficiles à interpréter.
Pajot *et al.*	1986	La leishmaniose en Guyane française. VI: Fluctuations saisonnières de la densité et du taux d'infection naturelle de *Lutzomyia* (*Nyssomyia*) *umbratilis* Ward et Fraiha, 1977 en forêt dégradée	Suivi de la population de phlébotomes en forêt perturbée pendant deux ans (dans la zone de Montsinéry). *Lu. umbratilis* est de loin le phlébotome le plus abondant mais d'autres espèces sont retrouvées par capture sur homme. *Lu. umbratilis* est plus abondant en saison des pluies, il n'y a pas de pic au mois de Novembre. Une variation d'abondance d'un facteur 3 est constatée entre les deux années. Le risque de transmission est maintenu toute la première année avec un pic en décembre mais est très faible sur la deuxième année.
Alexandre *et al.*	1987	La leishmaniose en Guyane française. VII: Caractéristiques structurales de quelques sites de contamination humaine en forêt	Les auteurs étudient les caractéristiques de 7 sites probables de contamination proches de Cayenne. Les sites sont tous situés dans des zones à forte pluviométrie et de forêt dense mais dégradée. La dégradation forestière apparait aux auteurs comme un facteur favorisant à la fois la multiplication des phlébotomes et la fréquentation humaine

Tableau 4 (suite): Synthèse (non-exhaustive) des études éco-épidémiologiques des leishmanioses réalisées en Guyane depuis le début des années 80.

Auteurs	**Année**	**Titre**	**Résumé**
Dedet *et al.*	1989	Isolation of *Leishmania* species from wild mammals in French Guiana	Recherche de *Leishmania* chez 486 mammifères appartenant à 20 espèces. *L. guyanensis* est détectée chez 11/31 *Choloepus didactylus*, 2/122 *Didelphis marsupialis* et 2/89 *Proechimys* sp. *L. amazonensis* est détectée chez 3/89 *Proechimys* sp.
Rotureau	2006	Ecology of the *Leishmania* species in the Guianan ecoregion complex	L'auteur établi une revue des connaissances écologiques des espèces *Leishmania* présentes dans la région
Rotureau *et al.*	2006	Diversity and ecology of sand flies (Diptera: Psychodidae: Phlebotominae) in coastal French Guiana	Des phlébotomes sont capturés à l'aide de piège CDC et identifiés dans des sites représentant 5 écotopes différents. 46 espèces de phlébotomes sont identifiées parmi 817 spécimens. La diversité la plus grande est retrouvée dans des sites de forêt primaire. L'abondance la plus élevée est retrouvée dans une zone de plantation, et la plus faible en zone urbaine à Cayenne. Globalement, *Lu. umbratillis* est l'espèce majoritaire. Des flagellés sont retrouvés dans 3/458 femelles disséquées, mais l'identification des parasites n'a pas pu être confirmée.
Rotureau *et al.*	2006	Absence of *Leishmania* in Guianan Bats	Après recherche par PCR chez 216 chauves-souris représentant 29 espèces, aucune *Leishmania* n'est détectée.
Rotureau	2006	Are New World leishmaniases becoming anthroponoses?	En regard de la plasticité épidémiologique des *Leishmania* et des changements environnementaux induits par l'homme, l'auteur discute de la possibilité de l'adaptation de ces parasites aux milieux anthropisés. Certaines espèces, comme *L. guyanensis*, semblent attachées à un écotope spécifique et pourrait disparaitre avec l'anthropisation des milieux tandis que d'autres, comme *L. braziliensis* ou *L. amazonensis*, pourrait s'adapter et devenir des zoonoses synanthropiques
Fouque *et al.*	2007	Phlebotomine sand flies (Diptera: Psychodidae) associated with changing patterns in the transmission of the human cutaneous leishmaniasis in French Guiana	2245 phlébotomes appartenant à 38 espèces ont été collectés dans différents sites à l'aide de pièges CDC. Les espèces les plus abondantes sont: *Lu. ininii*, *Lu. s. maripaensis* et *Lu. flaviscutellata*. Parmi 534 femelles disséquées pour la recherche de *Leishmania*, 5 sont trouvée positives. *Lu. flaviscutellata* a été trouvé infectée par *L. guyanensis*, *Lu. s. maripaensis* par *L. naiffi* et *Lu. ininii* par une espèce de *Leishmania* non-identifiée.

C. IMPACT DES ACTIVITÉS HUMAINES

Bien que l'imperfection des systèmes de surveillance dans la plupart des pays endémiques ne permette pas de disposer de données épidémiologiques précises, la leishmaniose est régulièrement qualifiée de maladie émergente ou réémergente dans la littérature. Une augmentation frappante de son incidence ou de sa distribution géographique a effectivement été observée dans certaines régions au cours des dernières décennies (au Brésil en particulier, mais également dans d'autres localités d'Amérique du Sud, d'Afrique et du Moyen Orient; Desjeux 2001, 2004). Selon les derniers chiffres de l'OMS, le nombre de cas déclarés de LC et de LV a augmenté à l'échelle mondiale depuis 2000, mais une diminution est observée sur les dernières années (WHO 2016, **Figure 24**). Ces tendances recouvrent toutefois une grande hétérogénéité spatiale, ainsi que la coexistence de facteurs explicatifs distincts. Dans la majorité des cas, ce sont des facteurs socio-économiques et démographiques qui sont en cause, dont les composantes majeures sont:

• les mouvements de populations dues aux sécheresses, aux famines, aux guerres, ou à la recherche de travail et qui provoquent la mise en contact d'individus non-immunisés avec le parasite. Un des exemples les plus dramatiques est l'épidémie de LV qui a eu lieu dans les années 1980 au Soudan, lorsque la guerre civile déplaça des populations dans des zones endémiques, faisant près de 100 000 morts selon Médecins Sans-Frontières (Seaman *et al.* 1996). Les frontières peuvent rassembler des populations migrantes d'origines diverses et constituent donc également des zones à risque: une épidémie de LC dans un camp de réfugiés Afghans au Pakistan aurait provoqué des lésions actives chez près de 40% des habitants (Rowland *et al.* 1999). Les opérations militaires ou le tourisme en zone d'endémie sont également responsables d'un nombre grandissant de cas de leishmaniose (Shaw 2007).

• l'urbanisation massive et dans de mauvaises conditions (typiquement, l'installation de bidonvilles). En effet, l'accumulation de matière organique due à la non gestion des déchets fournissent des sites de reproduction pour certaines espèces vectrices (*Lu. longipalpis* et *Lu. cruzi* par exemple) dans un environnement densément peuplé. L'exemple le plus frappant est celui du Brésil qui a connu, dans la seconde moitié du 20ème siècle, un fort exode rural vers les grandes villes. Cette « favelisation » s'est accompagnée, entre autres conséquences négatives, d'une propagation et d'une forte augmentation d'incidence de la LV et de la LC dans le pays (Harhay *et al.* 2011; Maroli *et al.* 2013). Des phénomènes similaires ont eu lieu dans d'autres pays d'Amérique Latine ainsi qu'au Moyen-Orient. Dans les banlieues pauvres de Kaboul,

par exemple, une épidémie de LC aurait touché près de 300 000 personnes à la fin des années 1990 (Hewitt *et al.* 1998).

• l'expansion de l'habitat humain proche de milieux sauvages où la leishmaniose circule de manière enzootique. Dans la ville de Manaus au Brésil, la construction de banlieues en bordure de forêt a provoqué la mise en contact de la population avec *L. guyanensis* (Barrett & Senra 1989). Dans la région andine, le manque de terres agricoles en montagne a conduit à une vague d'émigration vers les zones de forêt tropicale où la LC est enzootique, causant des épidémies importantes (WHO Expert Committee 2010).

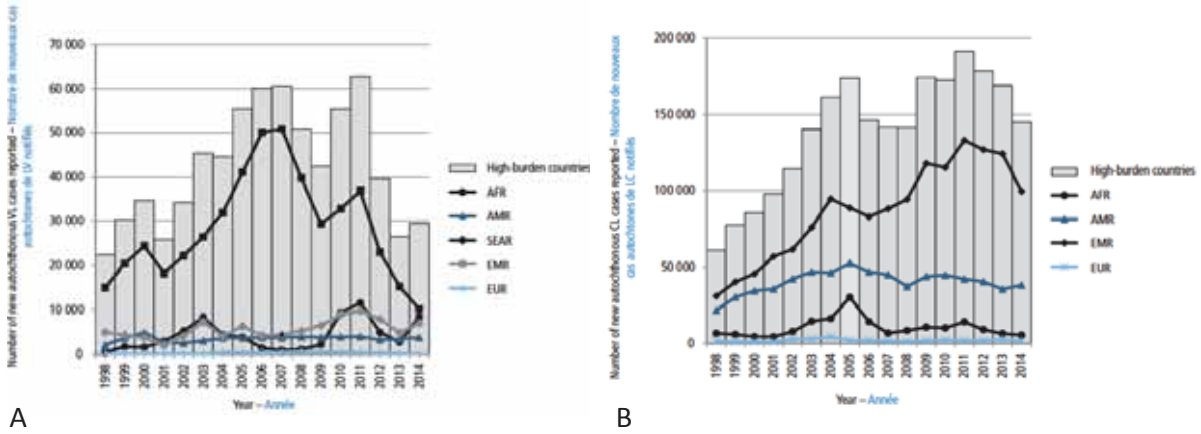
La pauvreté est généralement associée à toutes ces circonstances. Celle-ci s'accompagne de difficultés d'accès aux soins et de faiblesses physiologiques pour les populations concernée, ce qui accentue l'impact clinique de la maladie (Cerf *et al.* 1987).

Par ailleurs, la grande adaptabilité de certaines espèces d'hôtes et de vecteurs aux perturbations anthropiques et la possibilité d'une implication de la faune domestique dans le cycle de transmission, ont permis la « péridomestication » de la LC zoonotique dans certaines zones où il avait été envisagé que la déforestation puisse conduire à son éradication (Campbell-Lendrum *et al.* 2001). Au Venezuela, l'invasion intra-domiciliaire de certaines espèces de phlébotomes semble avoir contribué à une augmentation des cas de LC (Feliciangeli 1997; Desjeux 2001). Au Brésil, la présence de foyers de LC due à *L. braziliensis* transmis par *Lu. whitmani* ou *Lu. intermedia* a été rapportée dans des habitats colonisés depuis longtemps et où peu de zones forestières subsistent (Tolezano & Tolezano 1994; Brandão-Filho *et al.* 1999). De plus, les pratiques agriculturales (plantations, irrigation) peuvent parfois favoriser la prolifération d'hôtes et de vecteurs de leishmanioses (Desjeux 2001; Maroli *et al.* 2013).

L'épidémie de sida qui a débuté dans les années 1980 a également provoqué une recrudescence des leishmanioses en Europe et dans tous les autres continents où elles sévissent (Alvar *et al.* 2008). Du fait qu'elles ciblent les mêmes types cellulaires, les deux maladies ont un effet synergétique: le sida est associé à une augmentation extrême du risque de contracter la LV dans les zones d'endémicité et la LV accélère la progression clinique du sida.

La température, la pluviométrie et l'humidité sont des paramètres importants de l'écologie des phlébotomes. De petits changements de températures peuvent également avoir de sérieux effets sur le développement intravectoriel des *Leishmania* (Ready *et al.* 2010; WHO Expert Committee 2010). Des inquiétudes existent donc quant à une éventuelle expansion géographique des leishmanioses sous l'effet des changements climatiques.

Plusieurs modèles prédisent une progression vers des zones tempérées d'Europe et d'Amérique (González *et al.* 2010; Fischer *et al.* 2011) et des cas autochtones ont effectivement été rapportés récemment dans des zones où la maladie était inconnue jusqu'alors (dans le Nord de l'Italie, en Allemagne et dans les Pyrénées française; Maroli *et al.* 2008; Naucke *et al.* 2008; Dereure *et al.* 2009). L'existence d'un effet des changements climatiques sur l'épidémiologie de la leishmaniose en Europe n'est toutefois pas établi (Ready & others 2010).



A B

Figure 24: Evolution du nombre de cas annuels de LV **(A)** et de LC **(B)** rapportés dans les pays les plus touchés et par région OMS depuis 1998 (AFR=région africaine, AMR=région des Amériques, EMR=région de la Méditerranée orientale, EUR=région européenne, SEAR=région de l'Asie du Sud-Est, WPR= région du Pacifique occidental) (extrait de WHO 2016).

D. LEISHMANIOSES ET EFFET DE DILUTION

Dans certaines régions, les changements environnementaux d'origine humaine, tels que l'invasion des milieux sylvatiques et la déforestation, sont fréquemment incriminés comme des facteurs de réémergence des leishmanioses. Pourtant, l'impact épidémiologique de la perturbation des habitats naturels n'a quasiment jamais été étudié sous l'angle de l'effet de dilution pour ces maladies.

Plusieurs études ont démontré que certaines zones cultivées pouvaient constituer des environnements humides et riches en ressources, favorables aux vecteurs et hôtes réservoirs de certaines espèces de *Leishmania*, comme *L. amazonensis* dans le nouveau monde ou *L. major* dans l'ancien monde, contribuant ainsi à l'augmentation de la prévalence de ces parasites (Dedet *et al.* 1980; Ready *et al.* 1983; Ben Ammar *et al.* 1984; Walsh *et al.* 1993; Wasserberg *et al.* 2003). D'autres espèces de *Leishmania* comme *L. guyanensis* sont classiquement associées à un milieu sylvatique primaire, du fait de l'écologie de leurs hôtes et de leurs vecteurs. Certains travaux montrent dans ce cas une diminution du risque infectieux avec le niveau de perturbation ou l'éloignement de la forêt « intacte » (Le Pont & Pajot 1980;

Barrett & Senra 1989). Enfin, le potentiel zooprophylactique de la faune domestique non-compétente pour les *Leishmania*, tels que les lapins, les poules ou le bétail a également fait l'objet d'études dont les résultats ne sont pas unanimes (Alexander *et al.* 2002; Bern *et al.* 2005; Barnett *et al.* 2005; Chelbi *et al.* 2008).

Les conséquences de modifications plus fines des habitats et de la faune de vertébrés sur les communautés de phlébotomes et le cycle de transmission de ces parasites n'ont pas réellement été explorées. L'épidémiologie des leishmanioses zoonotiques implique pourtant des hôtes réservoirs sauvages ainsi que des vecteurs hématophages généralistes et représente donc un système sensible à la modification des communautés écologiques.

Pour plusieurs espèces de *Leishmania* (*L. major, L. mexicana, L. amazonensis,* et probablement *L. braziliensis* et *L. peruviana*), les hôtes réservoirs principaux sont des rongeurs ou des marsupiaux typiquement résilients aux perturbations anthropiques. Ceci peut expliquer, conjointement avec l'existence d'espèces vectrices capables des mêmes adaptations, la persistance de ces parasites dans des habitats perturbés voire péridomestiques. D'autres espèces de *Leishmania* circulent principalement chez des hôtes mammifères de plus grande taille et dont certains sont associés exclusivement au milieu sylvatique (*e.g*. xénarthres arboricoles pour *L. guyanensis et L. panamensis*, tatous pour *L. naiffi*). Bien que leur présence soit exclue dans des habitats totalement anthropisés, ces mammifères semblent présenter un certain niveau de résilience aux perturbations telles que la fragmentation forestière ou la chasse en comparaison d'autres espèces (Lopes & Ferrari 2000; Dalecky *et al.* 2002; Superina 2006; Michalski & Peres 2007; Stone *et al.* 2009; Catzeflis & Thoisy 2012). Leur persistance dans des communautés appauvries où d'autres hôtes moins compétents ne sont plus présents pourrait théoriquement conduire à amplifier la transmission des *Leishmania* dont ils sont porteurs.

La diversité d'hôtes secondaires occasionnellement incriminés dans l'épidémiologie des leishmanioses témoigne de la grande plasticité de ces parasites, et peut soulever des questions quant à leur adaptabilité à des communautés d'hôtes modifiées, même lorsque le réservoir principal habituel n'est plus présent (Rotureau 2006b). Par exemple, Arias *et al.* (1981) ont pu isoler *L. guyanensis* chez 20% (3/15) des opossums communs (*Didelphis marsupialis*, une espèce typiquement résiliente aux perturbations anthropiques) capturés dans une zone forestière considérée comme très perturbée, mais chez aucun des 27 capturés dans une zone proche considérée comme préservée.

De façon similaire, des interrogations existent quant à l'implication de la faune domestique et de rongeurs synanthropiques dans le maintien d'un cycle péridomestique pour

L. braziliensis et *L. peruviana* (Bonfante-Garrido *et al.* 1992; Reithinger & Davies 1999; Andrade *et al.* 2015). Réciproquement, bien que *L. chagasi (syn. L. infantum)* ait probablement été importé récemment sur le continent américain par des chiens infectés accompagnant les colons européens (Kuhls *et al.* 2011), et que le chien constitue toujours le réservoir principal en milieu anthropisé, le portage du parasite par d'autres mammifères sauvages permet l'existence de cycles sylvatiques (Lainson *et al.* 1990; Travi *et al.* 1994).

La transmission des *Leishmania* est également conditionnée par la présence des vecteurs. Si certaines espèces de phlébotomes sont particulièrement bien adaptées à des environnements hautement anthropisés comme *Lu. longipalpis*, d'autres, comme *Lu. umbratilis*, dépendent de la canopée et sont donc cantonnées au milieu forestier (Le Pont & Pajot 1980; Ready *et al.* 1983). Mais là encore, la plasticité épidémiologique des leishmanioses peut conduire à des situations inattendues. La capacité de certaines espèces de phlébotomes à permettre le développement de diverses espèces de *Leishmania*, indépendamment d'une complémentarité moléculaire très spécifique, a conduit à la notion d'espèces vectrices « permissives » (Volf & Peckova 2007). L'apparition de coadaptations vecteur-parasite peut permettre l'installation de cycles de transmission dans des zones où le vecteur habituel est absent. Le cas de l'association récente *L. chagasi (*syn. *L. infantum)-Lu. longipalpis* dans le nouveau monde constitue une mise en évidence frappante de ce phénomène, mais de nombreux autres exemples existent. Dans ce contexte, toutes modifications des communautés écologiques conduisant à la prédominance de certaines espèces de mammifères ou de phlébotomes peuvent potentiellement conduire à amplifier la prévalence d'une souche de *Leishmania* adaptée à ces hôtes.

Les recherches concernant l'effet de dilution souffrent d'un manque de données de terrain et de la restriction des études à un faible nombre d'exemples emblématiques. Les leishmanioses, du fait de la diversité des systèmes épidémiologiques qu'elles regroupent, représentent un modèle idéal pour étudier l'impact sanitaire des modifications des communautés écologiques par l'homme avec une perspective globale. Dans cette thèse, nous proposons donc d'explorer l'hypothèse de l'effet de dilution sur le modèle des leishmanioses zoonotiques. Ce travail vise à apporter de nouveaux éléments pour la compréhension des relations existantes entre biodiversité et maladies infectieuses, et à répondre à des questions d'intérêt sanitaires direct pour les régions ou ces maladies sont endémiques.

RÉFÉRENCES

Akhoundi M, Kuhls K, Cannet A *et al.* (2016) A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Neglected Tropical Diseases*, **10**, e0004349.

Akopyants NS, Kimblin N, Secundino N *et al.* (2009) Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science*, **324**, 265–268.

Alexander B, de Carvalho RL, McCallum H, Pereira MH (2002) Role of the domestic chicken (*Gallus gallus*) in the epidemiology of urban visceral leishmaniasis in Brazil. *Emerging Infectious Diseases*, **8**, 1480–1485.

Alexandre DY, Dedet J-P, Esterre P (1987) La leishmaniose en Guyane française. VII: caractéristiques structurales de quelques sites de contamination humaine en forêt. *Cahiers - ORSTOM. Entomologie médicale et parasitologie*, **25**, 101–109.

Allan BF, Keesing F, Ostfeld RS (2003) Effect of forest fragmentation on Lyme disease risk. *Conservation Biology*, **17**, 267–272.

Allan BF, Langerhans RB, Ryberg WA *et al.* (2009) Ecological correlates of risk and incidence of West Nile virus in the United States. *Oecologia*, **158**, 699–708.

de Almeida ME, Steurer FJ, Koru O *et al.* (2011) Identification of *Leishmania* spp. by molecular amplification and DNA sequencing analysis of a fragment of rRNA internal transcribed spacer 2. *Journal of Clinical Microbiology*, **49**, 3143–3149.

Altieri M, Nicholls C (2004) *Biodiversity and pest management in agroecosystems, Second Edition*. CRC Press.

Altizer S, Bartel R, Han BA (2011) Animal migration and infectious disease risk. *Science*, **331**, 296–302.

Altizer S, Ostfeld RS, Johnson PTJ, Kutz S, Harvell CD (2013) Climate change and infectious diseases: from evidence to a predictive framework. *Science*, **341**, 514–519.

Alvar J, Aparicio P, Aseffa A *et al.* (2008) The Relationship between leishmaniasis and AIDS: the Second 10 Years. *Clinical Microbiology Reviews*, **21**, 334–359.

Alvar J, Vélez ID, Bern C *et al.* (2012) Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE*, **7**, e35671.

Andrade MS, Courtenay O, Brito MEF *et al.* (2015) Infectiousness of Sylvatic and Synanthropic Small Rodents Implicates a Multi-host Reservoir of *Leishmania* (Viannia) *braziliensis*. *PLOS Negl Trop Dis*, **9**, e0004137.

Antonio L de F, Fagundes A, Oliveira RVC *et al.* (2014) Montenegro skin test and age of skin lesion as predictors of treatment failure in cutaneous leishmaniasis. *Revista do Instituto de Medicina Tropical de São Paulo*, **56**, 375–380.

Arango Duque G, Descoteaux A (2015) *Leishmania* survival in the macrophage: where the ends justify the means. *Current Opinion in Microbiology*, **26**, 32–40.

Aransay AM, Scoulica E, Chaniotis B, Tselentis Y (1999) Typing of sandflies from Greece and Cyprus by DNA polymorphism of 18S rRNA gene. *Insect Molecular Biology*, **8**, 179–184.

Arias JR, Naiff RD, Miles MA, de Souza AA (1981) The opossum, *Didelphis marsupialis* (Marsupialia: Didelphidae), as a reservoir host of *Leishmania braziliensis guyanensis* in the Amazon Basin of Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **75**, 537–541.

Arrivillaga JC, Feliciangeli MD (2001) *Lutzomyia pseudolongipalpis*: the first new species within the *longipalpis* (Diptera: Psychodidae: Phlebotominae) complex from La Rinconada, Curarigua, Lara State, Venezuela. *Journal of Medical Entomology*, **38**, 783–790.

Asato Y, Oshiro M, Myint CK *et al.* (2009) Phylogenic analysis of the genus *Leishmania* by cytochrome b gene sequencing. *Experimental Parasitology*, **121**, 352–361.

Bañuls AL, Bastien P, Pomares C *et al.* (2011) Clinical pleiomorphism in human leishmaniases, with special mention of asymptomatic infection. *Clinical Microbiology and Infection*, **17**, 1451–1461.

Bañuls AL, Hide M, Prugnolle F (2007) *Leishmania* and the leishmaniases: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. *Advances in Parasitology*, **64**, 1–109.

Banzet S (2000) Epidemies de leishmanioses cutanées chez les militaires en opération en Guyane Française. *Médecine tropicale*, **60**, 297–302.

Barker DC, Butcher J (1983) The use of DNA probes in the identification of leishmanias: discrimination between isolates of the *Leishmania mexicana* an *L. Braziliensis* complexes. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **77**, 285–297.

Barnett PG, Singh SP, Bern C, Hightower AW, Sundar S (2005) Virgin soil: the spread of visceral leishmaniasis into Uttar Pradesh, India. *The American Journal of Tropical Medicine and Hygiene*, **73**, 720–725.

Barrett R, Kuzawa CW, McDade T, Armelagos GJ (1998) Emerging and re-emerging infectious diseases: the third epidemiologic transition. *Annual Review of Anthropology*, **27**, 247–271.

Barrett TV, Senra MS (1989) Leishmaniasis in Manaus, Brazil. *Parasitology Today*, **5**, I989.

Basset D, Pratlong F, Ravel C *et al.* (2001) Les leishmanioses déclarées en France en 1999. *Bull. Épidémiol. Hebd*, **5**, 19–20.

Bates PA, Depaquit J, Galati E *et al.* (2015) Recent advances in phlebotomine sand fly research related to leishmaniasis control. *Parasites & Vectors*, **8**, 131.

Begon M (2008) Effects of host diversity on disease dynamics. *Infectious Disease Ecology: Effects of Ecosystems on Disease and of Disease on Ecosystems*, 12–29.

Ben Ammar R, Ben Ismail R, Helal H *et al.* (1984) Un nouveau foyer de leishmaniose cutanee de type rural dans la region de Sidi-Saad, Tunisie. *Bull. Soc. Fr. Parasitol*, **2**, 9–12.

Bern C, Hightower AW, Chowdhury R *et al.* (2005) Risk factors for Kala-Azar in Bangladesh. *Emerging Infectious Diseases*, **11**, 655–662.

Boer ML den, Alvar J, Davidson RN, Ritmeijer K, Balasegaram M (2009) Developments in the treatment of visceral leishmaniasis. *Expert Opinion on Emerging Drugs*, **14**, 395–410.

Bonfante-Garrido R, Meléndez E, Barroeta S *et al.* (1992) Cutaneous leishmaniasis in western Venezuela caused by infection with *Leishmania venezuelensis* and *L. braziliensis* variants. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **86**, 141–148.

Brandão-Filho SP, Campbell-Lendrum D, Brito MEF, Shaw JJ, Davies CR (1999) Epidemiological surveys confirm an increasing burden of cutaneous leishmaniasis in north-east Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **93**, 488–494.

Buffet PA, Rosenthal É, Gangneux J-P *et al.* (2011) Traitement des leishmanioses en France : proposition d'un référentiel consensuel. *La Presse Médicale*, **40**, 173–184.

Campbell LP, Luther C, Moo-Llanes D *et al.* (2015) Climate change influences on global distributions of dengue and chikungunya virus vectors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **370**, 20140135–20140135.

Campbell-Lendrum D, Dujardin J-P, Martinez E *et al.* (2001) Domestic and peridomestic transmission of American cutaneous leishmaniasis: changing epidemiological patterns present new control opportunities. *Memórias do Instituto Oswaldo Cruz*, **96**, 159–162.

Cantacessi C, Dantas-Torres F, Nolan MJ, Otranto D (2015) The past, present, and future of *Leishmania* genomics and transcriptomics. *Trends in Parasitology*.

Cardillo M, Mace GM, Gittleman JL *et al.* (2008) The predictability of extinction: biological and external correlates of decline in mammals. *Proceedings of the Royal Society of London B: Biological Sciences*, **275**, 1441–1448.

Cardinale BJ, Duffy JE, Gonzalez A *et al.* (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**, 59–67.

Carlson JC, Dyer LA, Omlin FX, Beier JC (2009) Diversity cascades and malaria vectors. *Journal of medical entomology*, **46**, 460.

Casanova C, Costa AI, Natal D (2005) Dispersal pattern of the sand fly *Lutzomyia neivai* (Diptera: Psychodidae) in a cutaneous leishmaniasis endemic rural area in Southeastern Brazil. *Memórias do Instituto Oswaldo Cruz*, **100**, 719–724.

Catzeflis F, Thoisy B de (2012) Xenarthrans in French Guiana: a brief overview of their distribution and conservation status. *Edentata*, **13**, 29–37.

Ceballos G, Ehrlich PR, Barnosky AD *et al.* (2015) Accelerated modern human–induced species losses: entering the sixth mass extinction. *Science Advances*, **1**, e1400253.

Cerf BJ, Jones TC, Badaro R *et al.* (1987) Malnutrition as a risk factor for severe visceral leishmaniasis. *The Journal of Infectious Diseases*, **156**, 1030–1033.

Chelbi I, Kaabi B, Derbali M *et al.* (2008) Zooprophylaxis: İmpact of breeding rabbits around houses on reducing the indoor abundance of *Phlebotomus papatasi*. *Vector-Borne and Zoonotic Diseases*, **8**, 741–748.

Chernin E, Perlstein JM (1971) Protection of snails against miracidia of *Schistosoma mansoni* by various aquatic invertebrates. *The Journal of Parasitology*, **57**, 217–219.

Chippaux J-P, Pajot F-X (1983) La leishmaniose en Guyane française. IV: note préliminaire sur les Phlébotomes des terriers. *Cahiers - ORSTOM. Entomologie médicale et parasitologie*, **21**, 149–154.

Christensen HA, Herrer A (1980) Panamanian *Lutzomyia* (Diptera: Psychodidae) host attraction profiles. *Journal of medical entomology*, **17**, 522–528.

Civitello DJ, Cohen J, Fatima H *et al.* (2015) Biodiversity inhibits parasites: broad evidence for the dilution effect. *Proceedings of the National Academy of Sciences*, 201506279.

Clay CA, Lehmer EM, Jeor SS, Dearing MD (2009a) Sin nombre virus and rodent species diversity: a test of the dilution and amplification hypotheses (M Baylis, Ed,). *PLoS ONE*, **4**, e6467.

Clay CA, Lehmer EM, St. Jeor S, Dearing MD (2009b) Testing mechanisms of the dilution effect: deer mice encounter rates, sin nombre virus prevalence and species diversity. *EcoHealth*, **6**, 250–259.

Cockburn A (1963) The evolution and eradication of infectious diseases. Johns Hopkins Univ. Press, Baltimore.

Colacicco-Mayhugh MG, Grieco JP, Putnam JL, Burkett DA, Coleman RE (2011) Impact of Phlebotomine sand flies on United States military operations at Tallil Air Base, Iraq: 5. Impact of weather on sand fly activity. *Journal of Medical Entomology*, **48**, 538–545.

Combes C, Moné H (1987) Possible mechanisms of the decoy effect in *Schistosoma mansoni* transmission. *International journal for parasitology*, **17**, 971–975.

Contreras Gutiérrez MA, Vivero RJ, Vélez ID, Porter CH, Uribe S (2014) DNA barcoding for the identification of sand fly species (Diptera, Psychodidae, Phlebotominae) in Colombia. *PLoS ONE*, **9**, e85496.

Cook RA, Karesh WB, Osofsky SA (2004) The Manhattan principles on "one world one health." *Buildling interdisciplinary bridges to health in a globalized world. New York, NY: Wildlife Conservation Society*.

Cooper N, Kamilar JM, Nunn CL (2012) Host longevity and parasite species richness in mammals. *PLOS ONE*, **7**, e42190.

Costa CHN, Stewart JM, Gomes RBB *et al.* (2002) Asymptomatic human carriers of *Leishmania chagasi*. *The American Journal of Tropical Medicine and Hygiene*, **66**, 334–337.

Cronin JP, Welsh ME, Dekkers MG, Abercrombie ST, Mitchell CE (2010) Host physiological phenotype explains pathogen reservoir potential. *Ecology Letters*, **13**, 1221–1232.

Cupolillo E, Medina-Acosta E, Noyes H *et al.* (2000) A revised classification for *Leishmania* and Endotrypanum. *Parasitology today*, **16**, 142–143.

Dalecky A, Chauvet S, Ringuet S *et al.* (2002) Large mammals on small islands: short term effects of forest fragmentation on the large mammal fauna in French Guiana. *Revue d'écologie. Supplément*, 145–164.

Darie H, Deniau M, Pratlong F *et al.* (1995) Cutaneous leishmaniasis of humans due to *Leishmania (Viannia) naiffi* outside Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **89**, 476–477.

Daszak P (2000) Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science*, **287**, 443–449.

Dedet J-P (1990) Cutaneous leishmaniasis in French Guiana: a review. *American journal of tropical medicine and hygiene*, **43**, 25–28.

Dedet JP (2009) Leishmanies, leishmanioses : biologie, clinique et thérapeutique. *EMC*.

Dedet JP, Desjeux P, Derouin F (1980) Ecologie d'un foyer de leishmaniose cutanée dans la region de Thies (Senegal, Afrique de l'Ouest). 4. Infestation spontanée et biologie de *Phlebotomus duboscqi* Neveu-Lemaire 1906. *Bulletin de la Societe de pathologie exotique et de ses filiales*.

Dedet JP, Desjeux P, Goyot P, Gosselin H (1984) Infestation naturelle de *Proechimys cuvieri* Petter 1978 (Rongeurs, Echimyidés) par *Leishmania mexicana amazonensis* Lainson et Shaw 1972 (Kinetoplastidé, Trypanosomatidé) en Guyane Française. *Comptes rendus des seances de l'Academie des Sciences. Serie III. Sciences de la Vie*.

Dedet JP, Gay F, Chatenay G (1989a) Isolation of *Leishmania* species from wild mammals in French Guiana. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **83**, 613–615.

Dedet JP, Pradinaud R, Desjeux P *et al.* (1985) Deux premiers cas de leishmaniose cutanée à *Leishmania mexicana amazonensis* en Guyane française. *Bulletin de la Société de pathologie exotique*, **78**, 64–70.

Dedet JP, Pradinaud R, Gay F (1989b) Epidemiological aspects of human cutaneous leishmaniasis in French Guiana. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **83**, 616–620.

Dedet JP, Pratlong F, Martini A, Lefebvre M, Dereure O (1994) *Leishmania (Viannia) brazilensis* en Guyane française: une réalité à prendre en considération. *Les Nouvelles dermatologiques*, **13**, 188–189.

Degrave W, Fernandes O, Campbell D *et al.* (1994) Use of molecular probes and PCR for detection and typing of *Leishmania* - a mini-review. *Memórias do Instituto Oswaldo Cruz*, **89**, 463–469.

Depaquit J (2014) Molecular systematics applied to Phlebotomine sandflies: review and perspectives. *Infection, Genetics and Evolution*, **28**, 744–756.

Dereure J, Vanwambeke SO, Malé P *et al.* (2009) The potential effects of global warming on changes in canine leishmaniasis in a focus outside the classical area of the disease in southern France. *Vector-Borne and Zoonotic Diseases*, **9**, 687–694.

Desjeux P (2001) The increase in risk factors for leishmaniasis worldwide. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **95**, 239–243.

Desjeux P (2004) Leishmaniasis: current situation and new perspectives. *Comparative Immunology, Microbiology and Infectious Diseases*, **27**, 305–318.

Desjeux P, Dedet JP (1989) Isoenzyme characterization of 112 *Leishmania* isolates from French Guiana. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **83**, 610–612.

Dick CW, Patterson BD (2007) Against all odds: explaining high host specificity in dispersal-prone parasites. *International Journal for Parasitology*, **37**, 871–876.

Dizney L, Dearing MD (2016) Behavioural differences: a link between biodiversity and pathogen transmission. *Animal Behaviour*, **111**, 341–347.

Dizney LJ, Ruedas LA (2009) Increased host species diversity and decreased prevalence of sin nombre virus. *Emerging Infectious Diseases*, **15**, 1012–1018.

Dobson A (2004) Population dynamics of pathogens with multiple host species. *The American Naturalist*, **164**, S64–S78.

Dobson A, Lafferty KD, Kuris AM, Hechinger RF, Jetz W (2008) Homage to Linnaeus: how many parasites? How many hosts? *Proceedings of the National Academy of Sciences*, **105**, 11482–11489.

Downing T, Imamura H, Decuypere S *et al.* (2011) Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Research*, **21**, 2143–2156.

Dunn RR, Davies TJ, Harris NC, Gavin MC (2010) Global drivers of human pathogen richness and prevalence. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20100340.

Dunn RR, Harris NC, Colwell RK, Koh LP, Sodhi NS (2009) The sixth mass coextinction: are most endangered species parasites and mutualists? *Proceedings of the Royal Society of London B: Biological Sciences*, **276**, 3037–3045.

El-Hossary S (2006) Morphological characteristics for sand fly taxonomy. *Research and Training Center on Vector of Diseases, Ain Shams University, Ain Shams*, 1–25.

Estes JA, Terborgh J, Brashares JS *et al.* (2011) Trophic downgrading of planet Earth. *Science*, **333**, 301–306.

Ezenwa VO, Godsey MS, King RJ, Guptill SC (2006) Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 109–117.

Fauci AS (2001) Infectious diseases: considerations for the 21st Century. *Clinical Infectious Diseases*, **32**, 675–685.

Feliciangeli MD (1997) Hourly activity of *Lutzomyia ovallesi* and *L. gomezi* (Diptera: Psychodidae), vectors of cutaneous leishmaniasis in northcentral Venezuela. *Journal of medical entomology*, **34**, 110–115.

Feliciangeli MD (2004) Natural breeding places of phlebotomine sandflies. *Medical and Veterinary Entomology*, **18**, 71–80.

Ferro C, Cárdenas E, Corredor D, Morales A, Munstermann LE (1998) Life cycle and fecundity analysis of *Lutzomyia shannoni* (Dyar) (Diptera: Psychodidae). *Memórias do Instituto Oswaldo Cruz*, **93**, 195–199.

Fischer D, Moeller P, Thomas SM, Naucke TJ, Beierkuhnlein C (2011) Combining climatic projections and dispersal ability: a method for estimating the responses of sand fly vector species to climate change. *PLOS Negl Trop Dis*, **5**, e1407.

Fisher RA, Corbet AS, Williams CB (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, **12**, 42–58.

Fisher MC, Henk DA, Briggs CJ *et al.* (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature*, **484**, 186–194.

Floch H (1943) Rapport sur le fonctionnement technique de l'Institut Pasteur de la Guyane pendant l'année 1943. *Pub Inst Pasteur Guyane française territ Inini*, **75**, 1–79.

Floch H (1954) *Leishmania tropica guyanensis* n.sp., pathogenic agent of Guyanese and Central American cutaneous leishmaniasis. *Bulletin de la Societe de pathologie exotique et de ses filiales*, **47**, 784–787.

Fouque F, Gaborit P, Issaly J *et al.* (2007) Phlebotomine sand flies (Diptera: Psychodidae) associated with changing patterns in the transmission of the human cutaneous leishmaniasis in French Guiana. *Memórias do Instituto Oswaldo Cruz*, **102**, 35–40.

Fraga J, Montalvo AM, Van der Auwera G *et al.* (2013) Evolution and species discrimination according to the *Leishmania* heat-shock protein 20 gene. *Infection, Genetics and Evolution*, **18**, 229–237.

Friggens MM, Beier P (2010) Anthropogenic disturbance and the risk of flea-borne disease transmission. *Oecologia*, **164**, 809–820.

Galati EAB (1995) Phylogenetic systematics of Phlebotominae (Diptera, Psychodidae) with emphasis on American groups. *Bol Dir Malariol San Amb*, **35**, 133–142.

Galati EAB (2014) Classificação, morfologia, terminologia e identificação de Adultos: Bioecologia e Identificação de Phlebotominae. In: *Flebotomíneos do Brasil* (eds Lainson R, Rangel EF), p. 367. Río de Janeiro.

Garin J-P, Piens MA, Pratlong F, Rioux J-A (1989) *Leishmania braziliensis* Vianna, 1911 en Guyane Française. *Bulletin de la Société de pathologie exotique*, **82**, 233–235.

GBD 2013 Mortality and Causes of Death Collaborators (2015) Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, **385**, 117–171.

Gentile B, Le Pont F, Pajot FX, Besnard R (1981) Dermal leishmaniasis in French Guiana: the sloth (*Choloepus didactylus*) as a reservoir host. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **75**, 612–613.

Geoffroy B, Dedet JP, Lebbe J, Esterre P, Trape JF (1986) Note sur les relations des vecteurs de leishmaniose avec les essences forestières en Guyane française. *Annales de Parasitologie Humaine et Comparée*, **61**, 483–490.

Gilbert L, Norman R, Laurenson KM, Reid HW, Hudson PJ (2001) Disease persistence and apparent competition in a three-host community: an empirical and analytical study of large-scale, wild populations. *Journal of Animal Ecology*, **70**, 1053–1061.

Gillespie PM, Beaumier CM, Strych U *et al.* (2016) Status of vaccine research and development of vaccines for leishmaniasis. *Vaccine*, **34**, 2992–2995.

González C, Wang O, Strutz SE *et al.* (2010) Climate change and risk of leishmaniasis in North America: predictions from ecological niche models of vector and reservoir species. *PLOS Negl Trop Dis*, **4**, e585.

Gonzalo JA, Alfonseca M, Muñoz F-F (2016) World Population: past, present and future. *World Scientific.*

Gottdenker NL, Calzada JE, Saldaña A, Carroll CR (2011) Association of anthropogenic land use change and increased abundance of the chagas disease vector *Rhodnius pallescens* in a rural landscape of Panama. *The American Journal of Tropical Medicine and Hygiene*, **84**, 70–77.

Gradoni L (2015) Canine *Leishmania* vaccines: still a long way to go. *Veterinary Parasitology*, **208**, 94–100.

Greger M (2007) The human/animal interface: emergence and resurgence of zoonotic infectious diseases. *Critical Reviews in Microbiology*, **33**, 243–299.

Guernaoui S, Boumezzough A (2009) Habitat preferences of phlebotomine sand flies (Diptera: Psychodidae) in Southwestern Morocco. *Journal of Medical Entomology*, **46**, 1187–1194.

Guernier V, Hochberg ME, Guégan J-F (2004) Ecology drives the worldwide distribution of human diseases. *PLOS Biol*, **2**, e141.

Hammond DS, Gond V, de Thoisy B, Forget P-M, DeDijn BPE (2007) Causes and consequences of a tropical forest gold rush in the Guiana Shield, South America. *AMBIO: A Journal of the Human Environment*, **36**, 661–670.

Han BA, Schmidt JP, Bowden SE, Drake JM (2015) Rodent reservoirs of future zoonotic diseases. *Proceedings of the National Academy of Sciences*, **112**, 7039–7044.

Harhay MO, Olliaro PL, Costa DL, Costa CHN (2011) Urban parasitology: visceral leishmaniasis in Brazil. *Trends in Parasitology*, **27**, 403–409.

Harkins KM, Schwartz RS, Cartwright RA, Stone AC (2016) Phylogenomic reconstruction supports supercontinent origins for *Leishmania*. *Infection, Genetics and Evolution*, **38**, 101–109.

Harvell CD, Mitchell CE, Ward JR *et al.* (2002) Climate warming and disease risks for terrestrial and marine biota. *Science*, **296**, 2158–2162.

Hay SI, Rogers DJ, Randolph SE *et al.* (2002) Hot topic or hot air? Climate change and malaria resurgence in East African highlands. *Trends in Parasitology*, **18**, 530–534.

Henke SE, Bryant FC (1999) Effects of coyote removal on the faunal community in western Texas. *The Journal of wildlife management*, 1066–1081.

Hewitt S, Reyburn H, Ashford R, Rowland M (1998) Anthroponotic cutaneous leishmaniasis in Kabul, Afghanistan: vertical distribution of cases in apartment blocks. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **92**, 273–274.

Huang ZYX, van Langevelde F, Prins HHT, de Boer WF (2015) Dilution versus facilitation: impact of connectivity on disease risk in metapopulations. *Journal of Theoretical Biology*, **376**, 66–73.

Hurk AF van den, Ritchie SA, Mackenzie JS (2009) Ecology and geographical expansion of japanese encephalitis virus. *Annual Review of Entomology*, **54**, 17–35.

Isaac NJ, Cowlishaw G (2004) How species respond to multiple extinction threats. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **271**, 1135–1141.

Ivens AC, Peacock CS, Worthey EA *et al.* (2005) The genome of the Kinetoplastid parasite, *Leishmania major*. *Science (New York, N.Y.)*, **309**, 436–442.

James AM, Oliver JH (1990) Feeding and host preference of immature *Ixodes dammini*, *I. scapularis*, and *I. pacificus* (Acari: Ixodidae). *Journal of Medical Entomology*, **27**, 324–330.

Jamjoom MB, Ashford RW, Bates PA *et al.* (2004) *Leishmania donovani* is the only cause of visceral leishmaniasis in East Africa; previous descriptions of *L. infantum* and from this region are a consequence of convergent evolution in the isoenzyme data. *Parasitology*, **129**, 399–409.

Jaouadi K, Haouas N, Chaara D *et al.* (2013) Phlebotomine (Diptera, Psychodidae) bloodmeal sources in Tunisian cutaneous leishmaniasis Foci: could *Sergentomyia minuta*; which is not an exclusive herpetophilic species, be implicated in the transmission of pathogens? *Annals of the Entomological Society of America*, **106**, 79–85.

Johnson PTJ, Hartson RB, Larson DJ, Sutherland DR (2008) Diversity and disease: community structure drives parasite transmission and host fitness. *Ecology Letters*, **11**, 1017–1026.

Johnson PT., Lund PJ, Hartson RB, Yoshino TP (2009) Community diversity reduces *Schistosoma mansoni* transmission, host pathology and human infection risk. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1657–1663.

Johnson NPAS, Mueller J (2002) Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bulletin of the History of Medicine*, **76**, 105–115.

Johnson PTJ, Ostfeld RS, Keesing F (2015) Frontiers in research on biodiversity and disease (E Seabloom, Ed,). *Ecology Letters*, ***18***, 1119-1133.

Johnson PTJ, Preston DL, Hoverman JT, Richgels KLD (2013) Biodiversity decreases disease through predictable changes in host community competence. *Nature*, **494**, 230–233.

Johnson PTJ, Rohr JR, Hoverman JT *et al.* (2012) Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. *Ecology Letters*, **15**, 235–242.

Johnson PTJ, Wood CL, Joseph MB *et al.* (2016) Habitat heterogeneity drives the host-diversity-begets-parasite-diversity relationship: evidence from experimental and field studies (D Ebert, Ed,). *Ecology Letters*, **19**, 752–761.

Jones BA, Grace D, Kock R *et al.* (2013) Zoonosis emergence linked to agricultural intensification and environmental change. *Proceedings of the National Academy of Sciences*, **110**, 8399–8404.

Jones KE, Patel NG, Levy MA *et al.* (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.

Kamhawi S (2000) The biological and immunomodulatory properties of sand fly saliva and its role in the establishment of *Leishmania* infections. *Microbes and Infection*, **2**, 1765–1773.

Kasap OE, Alten B (2006) Comparative demography of the sand fly *Phlebotomus papatasi* (Diptera: Psychodidae) at constant temperatures. *Journal of Vector Ecology*, **31**, 378–385.

Keeling MJ, Rohani P (2008) Modeling infectious diseases in humans and animals. Princeton University Press.

Keesing F, Belden LK, Daszak P *et al.* (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, **468**, 647–652.

Keesing F, Holt RD, Ostfeld RS (2006) Effects of species diversity on disease risk. *Ecology Letters*, **9**, 485–498.

Keesing F, Ostfeld RS (2015) Is biodiversity good for your health? *Science*, **349**, 234–235.

Killick-Kendrick R (1990) Phlebotomine vectors of the leishmaniases: a review. *Medical and Veterinary Entomology*, **4**, 1–24.

Killick-Kendrick R (1999) The biology and control of phlebotomine sand flies. *Clinics in dermatology*, **17**, 279–289.

Killick-Kendrick R, Killick-Kendrick M (1987) The laboratory colonization of *Phlebotomus ariasi* (Diptera: Psychodidae). *Annales de parasitologie humaine et comparée*, **62**, 354–356.

Killick-Kendrick R, Lainson R, Rioux JA, Saf'janova VM (1986) The taxonomy of *Leishmania*-like parasites of reptiles. *JA Rioux, Leishmania, Taxonomie et Phylogenèse. Application Éco-epidemiologiques*.

Kilpatrick AM, Altizer S (2010) Disease ecology. *Nature Education Knowledge*, **1**, 13.

Kilpatrick AM, Randolph SE (2012) Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *The Lancet*, **380**, 1946–1955.

Koarashi Y, Cáceres AG, Saca FMZ *et al.* (2016) Identification of causative *Leishmania* species in Giemsa-stained smears prepared from patients with cutaneous leishmaniasis in Peru using PCR-RFLP. *Acta Tropica*, **158**, 83–87.

Koch R (1876) Untersuchungen über Bacterien. V. Die Aetiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des *Bacillus Anthracis*. *Beiträge zur Biologie der Pflanzen*, **2**.

Krüger A, Strüven L, Post RJ, Faulde M (2011) The sandflies (Diptera: Psychodidae, Phlebotominae) in military camps in northern Afghanistan (2007–2009), as identified by morphology and DNA "barcoding." *Annals of Tropical Medicine & Parasitology*, **105**, 163–176.

Kuhls K, Alam MZ, Cupolillo E *et al.* (2011) Comparative microsatellite typing of New World *Leishmania infantum* reveals low heterogeneity among populations and its recent Old World Origin. *PLOS Negl Trop Dis*, **5**, e1155.

Lainson R, Dye C, Shaw JJ *et al.* (1990) Amazonian visceral leishmaniasis-distribution of the vector *Lutzomyia longipalpis* (Lutz & Neiva) in relation to the fox *cerdocyon thous* (Linn) and the efficiency of this reservoir host as a source of infection. *Memórias do Instituto Oswaldo Cruz*, **85**, 135–137.

Lainson R, Shaw JJ (1968) Leishmaniasis in Brazil: I. Observations on enzootic rodent leishmaniasis—incrimination of *Lutzomyia flaviscutellata* as the vector in the lower amazonian basin. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **62**, 396–405.

Lainson R, Shaw JJ (1987) Evolution, classification and geographical distribution. In: *The Leishmaniases in Biology and Medicine*, Academic Press, London.

Lainson R, Shaw JJ (2010) New world leishmaniasis. *Topley and Wilson's microbiology and microbial infections*.

Laurenson MK, Norman RA, Gilbert L, Reid HW, Hudson PJ (2003) Identifying disease reservoirs in complex systems: mountain hares as reservoirs of ticks and louping-ill virus, pathogens of red grouse. *Journal of Animal Ecology*, **72**, 177–185.

Lawyer P, Young D (1991) Diapause and quiescence in *Lutzomyia diabolica* (Diptera:Psychodidae). *Parassitologia*, **33 Suppl**, 353–360.

Le Pont F (1982) La leishmaniose en Guyane française: 2. Fluctuations saisonnières d'abondance et du taux d'infection naturelle de *Lutzomyia (Nyssomyia) umbratilis* Ward et Fraiha, 1977. *Cahiers ORSTOM. Série Entomologie Médicale et Parasitologie*, **20**, 269–277.

Le Pont F, Pajot F-X (1980) Leishmaniose en Guyane Francaise. 1. Etude de l'écologie et du taux d'infection naturelle du vecteur *Lutzomyia* (*Nyssomyia*) *umbratilis* Ward et Fraiha, 1977 en saison sèche. Considérations épidemiologiques. *Cahiers ORSTOM: entomologie medicale et parasitologie*, **18**, 359–382.

Le Pont F, Pajot F-X (1981) La leishmaniose en Guyane française: 2. Modalités de la transmission dans un village forestier: Cacao.

Le Pont R, Pajot F-X, Reguer R (1980) Preliminary observations on the silvatic cycle of leishmaniasis in French Guiana. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **74**, 133.

Levi T, Massey AL, Holt RD *et al.* (2016) Does biodiversity protect humans against infectious disease? Comment. *Ecology*, **97**, 536–542.

Lewis D, Young DG, Fairchild GB, Minter DM (1977) Proposals for a stable classification of the phlebotomine sandflies (Diptera: Psychodidae). *Systematic Entomology*, **2**, 319–332.

Lightburn E, Meynard JB, Morand J *et al.* (2001) Surveillance épidémiologique des leishmanioses tégumentaires en Guyane Synthèse de données militaires sur 10 ans. *Méd Trop*, **62**, 545–553.

Lindgren E, Gustafson R (2001) Tick-borne encephalitis in Sweden and climate change. *The Lancet*, **358**, 16–18.

Lindgren E, Tälleklint L, Polfeldt T (2000) Impact of climatic change on the northern latitude limit and population density of the disease-transmitting European tick *Ixodes ricinus*. *Environmental Health Perspectives*, **108**, 119–123.

Llanes A, Restrepo CM, Del Vecchio G, Anguizola FJ, Lleonart R (2015) The genome of *Leishmania panamensis*: insights into genomics of the *L. (Viannia)* subgenus. *Scientific Reports*, **5**, 8550.

Locatelli S, Peeters M (2012) Cross-species transmission of simian retroviruses: how and why they could lead to the emergence of new diseases in the human population. *AIDS*, **26**, 659–673.

LoGiudice K, Ostfeld RS, Schmidt KA, Keesing F (2003) The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. *Proceedings of the National Academy of Sciences*, **100**, 567–571.

Lopes MA, Ferrari SF (2000) Effects of human colonization on the abundance and diversity of mammals in eastern Brazilian Amazonia. *Conservation Biology*, **14**, 1658–1665.

Loreau M, Naeem S, Inchausti P *et al.* (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science*, **294**, 804–808.

Lovell WG (1992) "Heavy shadows and black night": disease and depopulation in colonial Spanish America. *Annals of the Association of American Geographers*, **82**, 426–443.

Lun Z-R, Wu M-S, Chen Y-F *et al.* (2015) Visceral leishmaniasis in China: an endemic disease under control. *Clinical Microbiology Reviews*, **28**, 987–1004.

MacArthur RH, Diamond JM, Karr JR (1972) Density Compensation in Island Faunas. *Ecology*, **53**, 330–342.

Macdonald G (1956) Epidemiological basis of malaria control. *Bulletin of the World Health Organization*, **15**, 613–626.

Marco JD, Uezato H, Mimori T *et al.* (2006) Are cytochrome B gene sequencing and polymorphism-specific polymerase chain reaction as reliable as multilocus enzyme electrophoresis for identifying *Leishmania* spp. from Argentina? *The American journal of tropical medicine and hygiene*, **75**, 256–260.

Marfurt J, Nasereddin A, Niederwieser I *et al.* (2003) Identification and differentiation of *Leishmania* species in clinical samples by PCR amplification of the miniexon sequence and subsequent restriction fragment length polymorphism analysis. *Journal of Clinical Microbiology*, **41**, 3147–3153.

Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L (2013) Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Medical and Veterinary Entomology*, **27**, 123–147.

Maroli M, Rossi L, Baldelli R *et al.* (2008) The northward spread of leishmaniasis in Italy: evidence from retrospective and ongoing studies on the canine reservoir and phlebotomine vectors. *Tropical Medicine & International Health*, **13**, 256–264.

Marques MJ, Volpini ÂC, Machado-Coelho GLL *et al.* (2006) Comparison of polymerase chain reaction with other laboratory methods for the diagnosis of American cutaneous leishmaniasis: diagnosis of cutaneous leishmaniasis by polymerase chain reaction. *Diagnostic Microbiology and Infectious Disease*, **54**, 37–43.

Martin LB, Hasselquist D, Wikelski M (2006) Investment in immune defense is linked to pace of life in house sparrows. *Oecologia*, **147**, 565–575.

Martin-Blondel G, Iriart X, El Baidouri F *et al.* (2015) Outbreak of *Leishmania braziliensis* cutaneous leishmaniasis, Saül, French Guiana. *Emerging Infectious Diseases*, **21**, 892–894.

Matuschka F-R, Heiler M, Eiffert H *et al.* (1993) Diversionary role of hoofed game in the transmission of Lyme disease spirochetes. *The American journal of tropical medicine and hygiene*, **48**, 693–699.

Matuschka F-R, Spielman A (1992) Loss of Lyme disease spirochetes from *Ixodes ricinus* ticks feeding on European blackbirds. *Experimental Parasitology*, **74**, 151–158.

McMichael AJ (2004) Environmental and social influences on emerging infectious diseases: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 1049–1058.

McNeill W (1977) Plagues and People. *Anchor Books*.

Michalski F, Peres CA (2007) Disturbance-mediated mammal persistence and abundance-area relationships in Amazonian forest fragments. *Conservation Biology*, ***21***, 1626-1640.

Mihaljevic JR, Joseph MB, Orlofske SA, Paull SH (2014) The scaling of host density with richness affects the direction, shape, and detectability of diversity-disease relationships. *PLOS ONE*, **9**, e97812.

Miller E, Huppert A (2013) The effects of host diversity on vector-borne disease: the conditions under which diversity will amplify or dilute the disease risk. *PLOS ONE*, **8**, e80279.

Minter LM, Yu T, Florin DA *et al.* (2013) Molecular Identification of Sand Flies (Diptera: Psychodidae) in Eastern North America by Using PCR-RFLP. *Journal of Medical Entomology*, **50**, 920–924.

Mohebali M, Hajjaran H, Hamzavi Y *et al.* (2005) Epidemiological aspects of canine visceral leishmaniosis in the Islamic Republic of Iran. *Veterinary Parasitology*, **129**, 243–251.

Monge-Maillo B, López-Vélez R (2015) Miltefosine for visceral and cutaneous leishmaniasis: drug characteristics and evidence-based treatment recommendations. *Clinical Infectious Diseases*, **60**, 1398–1404.

Montenegro J (1926) A cutis-reação na leishmaniose. *An. Fac. Med. S. Paulo*, **1**, 323–330.

Morand S, Jittapalapong S, Suputtamongkol Y, Abdullah MT, Huan TB (2014) Infectious diseases and their outbreaks in Asia-Pacific: biodiversity and its regulation loss matter. *PLOS ONE*, **9**, e90032.

Morens DM, Folkers GK, Fauci AS (2004) The challenge of emerging and re-emerging infectious diseases. *Nature*, **430**, 242–249.

Morrison AC, Ferro C, Morales A, Tesh RB, Wilson ML (1993) Dispersal of the sand fly *Lutzomyia longipalpis* (Diptera: Psychodidae) at an endemic focus of visceral leishmaniasis in Colombia. *Journal of medical entomology*, **30**, 427–435.

Morse S (1995) Factors in the emergence of infectious diseases. *Emerging Infectious Diseases*, **1**, 7–15.

Muller G, Schlein Y (2004) Nectar and honeydew feeding of *Phlebotomus papatasi* in a focus of *Leishmania major* in Neot Hakikar oasis. *Journal of vector ecology*, **29**, 154–158.

Myers N, Myers NH, Wachtel C *et al.* (1985) The primary source: tropical forests and our future. *National Academy of Sciences*, Washington, DC (EUA).

Naucke TJ, Menn B, Massberg D, Lorentz S (2008) Sandflies and leishmaniasis in Germany. *Parasitology research*, **103**, 65–68.

Norman R, Bowers RG, Begon M, Hudson PJ (1999) Persistence of tick-borne virus in the presence of multiple host species: tick reservoirs and parasite mediated competition. *Journal of Theoretical Biology*, **200**, 111–118.

Nupp TE, Swihart RK (1998) Effects of forest fragmentation on population attributes of white-footed mice and Eastern Chipmunks. *Journal of Mammalogy*, **79**, 1234–1243.

Nzelu CO, Cáceres AG, Guerrero-Quincho S *et al.* (2016) A rapid molecular diagnosis of cutaneous leishmaniasis by colorimetric malachite green-loop-mediated isothermal amplification (LAMP) combined with an FTA card as a direct sampling tool. *Acta Tropica*, **153**, 116–119.

Ogden NH, Lindsay LR, Morshed M, Sockett PN, Artsob H (2009) The emergence of Lyme disease in Canada. *Canadian Medical Association Journal*, **180**, 1221–1224.

Ogden NH, Maarouf A, Barker IK *et al.* (2006) Climate change and the potential for range expansion of the Lyme disease vector *Ixodes scapularis* in Canada. *International Journal for Parasitology*, **36**, 63–70.

de Oliveira AG, Marassá AM, Consales CA *et al.* (2008) Observations on the feeding habits of *Lutzomyia longipalpis* (Lutz & Neiva, 1912) (Diptera: Psychodidae: Phlebotominae) in Campo Grande, an endemic area of visceral leishmaniasis in Mato Grosso do Sul, Brazil. *Acta Tropica*, **107**, 238–241.

Oliveira AG, Galati EAB, Fernandes CE, Dorval MEC, Brazil RP (2008) Seasonal variation of *Lutzomyia longipalpis* (Lutz & Neiva, 1912) (Diptera: Psychodidae: Phlebotominae) in endemic area of visceral leishmaniasis, Campo Grande, state of Mato Grosso do Sul, Brazil. *Acta Tropica*, **105**, 55–61.

OMS (2012) The global report for research on infectious diseases of poverty.

Orshan L, Elbaz S, Ben-Ari Y *et al.* (2016) Distribution and dispersal of *Phlebotomus papatasi* (Diptera: Psychodidae) in a zoonotic cutaneous leishmaniasis focus, the Northern Negev, Israel. *PLOS Negl Trop Dis*, **10**, e0004819.

Ostfeld RS, Keesing F (2000a) Biodiversity and disease risk: the case of Lyme disease. *Conservation Biology*, **14**, 722–728.

Ostfeld RS, Keesing F (2000b) Biodiversity series: the function of biodiversity in the ecology of vector-borne zoonotic diseases. *Canadian Journal of Zoology*, **78**, 2061–2078.

Ostfeld RS, Keesing F (2012) Effects of host diversity on infectious disease. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 157–182.

Ostfeld RS, Roy P, Haumaier W *et al.* (2004) Sand Fly (*Lutzomyia vexator*) (Diptera: Psychodidae) populations in upstate New York: abundance, microhabitat, and phenology. *Journal of Medical Entomology*, **41**, 774–778.

Packer C, Holt RD, Hudson PJ, Lafferty KD, Dobson AP (2003) Keeping the herds healthy and alert: implications of predator control for infectious disease. *Ecology Letters*, **6**, 797–802.

Pajot F-X, Chippaux J-P, Geoffroy B, Dedet J-P (1986) La leishmaniose en Guyane française. VI: Fluctuations saisonnières de la densité et du taux d'infection naturelle de *Lutzomyia (Nyssomyia) umbratilis* Ward et Fraiha, 1977 en forêt dégradée. *Cahiers-ORSTOM. Entomologie médicale et parasitologie*, **24**, 191–198.

Park AW, Cleveland CA, Dallas TA, Corn JL (2015) Vector species richness increases haemorrhagic disease prevalence through functional diversity modulating the duration of seasonal transmission. *Parasitology*, 1–6.

Parker IM, Saunders M, Bontrager M *et al.* (2015) Phylogenetic structure and host abundance drive disease pressure in communities. *Nature*, **520**, 542–544.

Patz JA, Daszak P, Tabor GM *et al.* (2004) Unhealthy landscapes: policy recommendations on land use change and infectious disease emergence. *Environmental Health Perspectives*, **112**, 1092–1098.

Peacock CS, Seeger K, Harris D *et al.* (2007) Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nature Genetics*, **39**, 839–847.

Peiris JSM, Yuen KY, Osterhaus ADME, Stöhr K (2003) The severe acute respiratory syndrome. *New England Journal of Medicine*, **349**, 2431–2441.

Peres CA, Dolman PM (2000) Density compensation in neotropical primate communities: evidence from 56 hunted and nonhunted Amazonian forests of varying productivity. *Oecologia*, **122**, 175–189.

Pigott DM, Bhatt S, Golding N *et al.* (2014) Global distribution maps of the leishmaniases. *eLife*, **3**, e02851.

Plowright RK, Eby P, Hudson PJ *et al.* (2014) Ecological dynamics of emerging bat virus spillover. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20142124–20142124.

Postigo JAR (2010) Leishmaniasis in the world health organization Eastern Mediterranean region. *International Journal of Antimicrobial Agents*, **36**, Supplement 1, S62–S65.

Poulin R, Morand S (2014) Parasite Biodiversity. *Smithsonian Institution*.

Preston FW (1948) The commonness and rarity of species. *Ecology*, **29**, 254–283.

Previtali MA, Ostfeld RS, Keesing F *et al.* (2012) Relationship between pace of life and immune responses in wild rodents. *Oikos*, **121**, 1483–1492.

Quaresma PF, Rêgo FD, Botelho HA *et al.* (2011) Wild, synanthropic and domestic hosts of *Leishmania* in an endemic area of cutaneous leishmaniasis in Minas Gerais State, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **105**, 579–585.

Rand PW, Lubelczyk C, Holman MS, Lacombe EH, Smith RP (2004) Abundance of *Ixodes scapularis* (Acari: Ixodidae) after the complete removal of deer from an isolated offshore island, endemic for Lyme disease. *Journal of Medical Entomology*, **41**, 779–784.

Randolph SE (2001) The shifting landscape of tick-borne zoonoses: tick-borne encephalitis and Lyme borreliosis in Europe. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **356**, 1045–1056.

Randolph SE, Dobson ADM (2012) Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology*, **139**, 847–863.

Rangel EF, Vilela ML (2008) *Lutzomyia longipalpis* (Diptera, Psychodidae, Phlebotominae) and urbanization of visceral leishmaniasis in Brazil. *Cadernos de Saúde Pública*, **24**, 2948–2952.

Read LK, Lukeš J, Hashimi H (2015) Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley Interdisciplinary Reviews: RNA*, n/a-n/a.

Ready PD (1979) Factors affecting egg oroduction of laboratory-bred *Lutzomyia longipalpis* (Diptera: Psychodidae). *Journal of Medical Entomology*, **16**, 413–423.

Ready PD (2013) Biology of phlebotomine sand flies as vectors of disease agents. *Annual Review of Entomology*, **58**, 227–250.

Ready PD, Lainson R, Shaw JJ (1983) Leishmaniasis in Brazil: XX. Prevalence of "enzootic rodent leishmaniasis"(*Leishmania mexicana amazonensis*), and apparent absence of "pian bois"(*Le. braziliensis guyanensis*), in plantations of introduced tree species and in other non-climax forests in eastern Amazonia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **77**, 775–785.

Ready PD, others (2010) Leishmaniasis emergence in Europe. *Euro surveill*, **15**, 19505.

Real F, Vidal RO, Carazzolle MF *et al.* (2013) The genome sequence of *Leishmania* (*Leishmania*) *amazonensis*: functional annotation and extended analysis of gene models. *DNA Research*, **20**, 567–581.

Reithinger R, Davies CR (1999) Is the domestic dog (*Canis familiaris*) a reservoir host of American cutaneous leishmaniasis? A critical review of the current evidence. *The American Journal of Tropical Medicine and Hygiene*, **61**, 530–541.

Rioux JA, Lanotte G, Serres E *et al.* (1990) Taxonomy of *Leishmania*. Use of isoenzymes. Suggestions for a new classification. *Ann Parasitol Hum Comp*, **65**, 111–125.

Rispail P, Léger N (1998) Numerical taxonomy of Old World Phlebotominae (Diptera: Psychodidae): 1. Considerations of morphological characters in the genus *Phlebotomus* Rondani and Berté 1840. *Memórias do Instituto Oswaldo Cruz*, **93**, 773–785.

Roche B, Rohani P, Dobson AP, Guégan J-F (2013) The impact of community organization on vector-borne pathogens. *The American Naturalist*, **181**, 1–11.

Rochlin I, Ninivaggi DV, Hutchinson ML, Farajollahi A (2013) Climate change and range expansion of the Asian tiger mosquito (*Aedes albopictus*) in Northeastern USA: implications for public health practitioners. *PLOS ONE*, **8**, e60874.

Rogers MB, Hilley JD, Dickens NJ *et al.* (2011) Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Research*, **21**, 2129–2142.

Rohr JR, Dobson AP, Johnson PTJ *et al.* (2011) Frontiers in climate change–disease research. *Trends in Ecology & Evolution*, **26**, 270–277.

Romano A, Inbar E, Debrabant A *et al.* (2014) Cross-species genetic exchange between visceral and cutaneous strains of *Leishmania* in the sand fly vector. *Proceedings of the National Academy of Sciences*, **111**, 16808–16813.

Romero-Ricardo L, Lastre-Meza N, Pérez-Doria A, Bejarano EE (2016) DNA barcoding to identify species of phlebotomine sand fly (Diptera: Psychodidae) in the mixed leishmaniasis focus of the Colombian Caribbean. *Acta Tropica*, **159**, 125–131.

Roscher C, Schumacher J, Foitzik O, Schulze E-D (2007) Resistance to rust fungi in *Lolium perenne* depends on within-species variation and performance of the host species in grasslands of different plant diversity. *Oecologia*, **153**, 173–183.

Rosenthal SR, Ostfeld RS, McGarvey ST, Lurie MN, Smith KF (2015) Redefining disease emergence to improve prioritization and macro-ecological analyses. *One Health*.

Rossi E, Bongiorno G, Ciolli E *et al.* (2008) Seasonal phenology, host-blood feeding preferences and natural *Leishmania* infection of *Phlebotomus perniciosus* (Diptera, Psychodidae) in a high-endemic focus of canine leishmaniasis in Rome province, Italy. *Acta Tropica*, **105**, 158–165.

Rottstock T, Joshi J, Kummer V, Fischer M (2014) Higher plant diversity promotes higher diversity of fungal pathogens, while it decreases pathogen infection per plant. *Ecology*, **95**, 1907–1917.

Rotureau B (2005a) Eco-épidémiologie des leishmanioses cutanées en Guyane française. Antilles-Guyane.

Rotureau B (2005b) Eco-épidémiologie des leishmanioses cutanées en Guyane française. Antilles-Guyane.

Rotureau B (2006a) Ecology of the *Leishmania* species in the Guianan ecoregion complex. *The American journal of tropical medicine and hygiene*, **74**, 81–96.

Rotureau B (2006b) Are New World leishmaniases becoming anthroponoses? *Medical Hypotheses*, **67**, 1235–1241.

Rotureau B, Catzeflis F, Carme B (2006a) Absence of *Leishmania* in Guianan Bats. *The American Journal of Tropical Medicine and Hygiene*, **74**, 318–321.

Rotureau B, Gaborit P, Issaly J *et al.* (2006b) Diversity and ecology of sand flies (Diptera: Psychodidae: Phlebotominae) in coastal French Guiana. *American Journal of Tropical Medicine and Hygiene*, **75**, 62.

Rougeron V, De Meeûs T, Bañuls A-L (2015) A primer for *Leishmania* population genetic studies. *Trends in Parasitology*, **31**, 52–59.

Rowland M, Munir A, Durrani N, Noyes H, Reyburn H (1999) An outbreak of cutaneous leishmaniasis in an Afghan refugee settlement in north-west Pakistan. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **93**, 133–136.

Rudolf VHW, Antonovics J (2005) Species coexistence and pathogens with frequency- dependent transmission. *The American Naturalist*, **166**, 112–118.

Ruedas LA, Salazar-Bravo J, Tinnin DS *et al.* (2004) Community ecology of small mammal populations in Panama following an outbreak of Hantavirus pulmonary syndrome. *Journal of Vector Ecology*, **29**, 177–191.

Ruiter CM de, Veer C van der, Leeflang MMG *et al.* (2014) Molecular tools for diagnosis of visceral leishmaniasis: systematic review and meta-analysis of diagnostic test accuracy. *Journal of Clinical Microbiology*, **52**, 3147–3155.

Ruiz-Moreno D, Vargas IS, Olson KE, Harrington LC (2012) Modeling dynamic introduction of Chikungunya virus in the United States. *PLOS Negl Trop Dis*, **6**, e1918.

Saldarriaga OA, Castellanos-Gonzalez A, Porrozzi R *et al.* (2016) An innovative field-applicable molecular test to diagnose cutaneous *Leishmania Viannia spp* . infections. *PLOS Negl Trop Dis*, **10**, e0004638.

Salkeld DJ, Padgett KA, Jones JH (2013) A meta-analysis suggesting that the relationship between biodiversity and risk of zoonotic pathogen transmission is idiosyncratic (K Lafferty, Ed,). *Ecology Letters*, **16**, 679–686.

Salkeld DJ, Padgett KA, Jones JH, Antolin MF (2015) Public health perspective on patterns of biodiversity and zoonotic disease. *Proceedings of the National Academy of Sciences*, 201517640.

Sang DK, Njeru WK, Ashford RW (1992) A possible animal reservoir for *Leishmania tropica sl.* in Kenya. *Annals of tropical medicine and parasitology*, **86**, 311.

Sax DF, Gaines SD (2003) Species diversity: from global decreases to local increases. *Trends in Ecology & Evolution*, **18**, 561–566.

Schlein Y, Jacobson RL (1999) Sugar meals and longevity of the sandfly *Phlebotomus papatasi* in an arid focus of *Leishmania major* in the Jordan Valley. *Medical and Veterinary Entomology*, **13**, 65–71.

Schlein Y, Muller G (1995) Assessment of plant tissue Feeding by sand flies (Diptera: Psychodidae) and mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology*, **32**, 882–887.

Schmidt KA, Ostfeld RS (2001) Biodiversity and the dilution effect in disease ecology. *Ecology*, **82**, 609–619.

Schönian G, Mauricio I, Cupolillo E (2010) Is it time to revise the nomenclature of *Leishmania* ? *Trends in Parasitology*, **26**, 466–469.

Schönian G, Schnur L, Fari ME *et al.* (2001) Genetic heterogeneity in the species *Leishmania tropica* revealed by different PCR-based methods. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **95**, 217–224.

Seaman J, Mercer AJ, Sondorp HE, Herwaldt BL (1996) Epidemic visceral leishmaniasis in Southern Sudan: treatment of severely debilitated patients under wartime conditions and with limited resources. *Annals of Internal Medicine*, **124**, 664–672.

Senghor MW, Niang AA, Depaquit J *et al.* (2016) Transmission of *Leishmania infantum* in the canine leishmaniasis focus of Mont-Rolland, Senegal: ecological, parasitological and molecular evidence for a possible role of *Sergentomyia* sand flies. *PLoS Negl Trop Dis*, **10**, e0004940.

Shapiro T, Klein V, Englund P (1999) Isolation of Kinetoplast DNA. In: *DNA Topoisomerase Protocols Methods in Molecular Biology™*. (eds Bjornsti M-A, Osheroff N), pp. 61–67. Humana Press.

Shaw JJ (2002) New World leishmaniasis: the ecology of leishmaniasis and the diversity of leishmanial species in Central and South America. In: *Leishmania*, pp. 11–31. Springer.

Shaw J (2007) The leishmaniases - survival and expansion in a changing world: a mini-review. *Memórias do Instituto Oswaldo Cruz*, **102**, 541–547.

Shaw JJ, Burger JF, others (1999) The relationship of sand fly ecology to the transmission of leishmaniasis in South America with particular reference to Brazil. *Contributions to the knowledge of Diptera: a collection of articles on Diptera commemorating the life and work of Graham B. Fairchild.*, 503–517.

Shaw JJ, Lainson R (1972) Leishmaniasis in Brazil: VI. Observations on the seasonal variations of *Lutzomyia flaviscutellata* in different types of forest and its relationship to enzootic rodent leishmaniasis (*Leishmania mexicana amazonensis*). *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **66**, 709–717.

Simon S, Nacher M, Carme B *et al.* (2017) Cutaneous leishmaniasis in French Guiana: revising epidemiology with PCR-RFLP. *Tropical Medicine and Health*, **45**, 5.

Simpson L (1987) The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. *Annual Review of Microbiology*, **41**, 363–380.

Sobrino R, Ferroglio E, Oleaga A *et al.* (2008) Characterization of widespread canine leishmaniasis among wild carnivores from Spain. *Veterinary Parasitology*, **155**, 198–203.

Springbett AJ, MacKenzie K, Woolliams JA, Bishop SC (2003) The contribution of genetic diversity to the spread of infectious diseases in livestock populations. *Genetics*, **165**, 1465–1474.

Steere AC, Coburn J, Glickstein L (2004) The emergence of Lyme disease. *Journal of Clinical Investigation*, **113**, 1093–1101.

Sterkers Y, Lachaud L, Bourgeois N *et al.* (2012) Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. *Molecular Microbiology*, **86**, 15–23.

Stone AI, Lima EM, Aguiar GFS *et al.* (2009) Non-volant mammalian diversity in fragments in extreme eastern Amazonia. *Biodiversity and Conservation*, **18**, 1685–1694.

Superina M (2006) The Biology of the *Xenarthra*. *Edentata*, **7**, 59.

Suzán G, Marcé E, Giermakowski JT *et al.* (2009) Experimental evidence for reduced rodent diversity causing increased Hantavirus prevalence (A Wilby, Ed,). *PLoS ONE*, **4**, e5461.

Svobodova M, Votypka J, Peckova J *et al.* (2006) Distinct transmission cycles of *Leishmania tropica* in 2 adjacent foci, Northern Israel. *Emerg Infect Dis*, **12**, 1860–1868.

Swaddle JP, Calos SE (2008) Increased avian diversity is associated with lower incidence of human West Nile infection: observation of the dilution effect (A Buckling, Ed,). *PLoS ONE*, **3**, e2488.

Swei A, Ostfeld RS, Lane RS, Briggs CJ (2011) Impact of the experimental removal of lizards on Lyme disease risk. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 2970–2978.

Taylor LH, Latham SM, Mark EJ (2001) Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **356**, 983–989.

Teixeira DE, Benchimol M, Rodrigues JCF *et al.* (2013) The cell biology of *Leishmania* : how to teach using animations. *PLOS Pathog*, **9**, e1003594.

Telfer S, Bown KJ, Sekules R *et al.* (2005) Disruption of a host-parasite system following the introduction of an exotic host species. *Parasitology*, **130**, 661–668.

Terayama Y, Kato H, Gomez EA *et al.* (2008) Molecular typing of sand fly species (Diptera, Psychodidae, Phlebotominae) from areas endemic for leishmaniasis in Ecuador by PCR-RFLP of 18S ribosomal RNA gene. *Journal of Veterinary Medical Science*, **70**, 907–913.

Tesh RB, Chaniotis BN, Aronson MD, Johnson KM, others (1971) Natural host preferences of Panamanian phlebotomine sandflies as determined by precipitin test. *American Journal of Tropical Medicine and Hygiene*, **20**, 150–156.

Thomaz-Soccol V, Lanotte G, Rioux JA *et al.* (1993) Monophyletic origin of the genus *Leishmania* Ross, 1903. *Annales de parasitologie humaine et comparée*, **68**, 107–107.

Thongsripong P, Green A, Kittayapong P *et al.* (2013) Mosquito vector diversity across habitats in Central Thailand endemic for dengue and other arthropod-borne diseases. *PLOS Negl Trop Dis*, **7**, e2507.

Tibayrenc M (2006) The species concept in parasites and other pathogens: a pragmatic approach? *Trends in Parasitology*, **22**, 66–70.

Tilman D, Fargione J, Wolff B *et al.* (2001) Forecasting agriculturally driven global environmental change. *Science*, **292**, 281–284.

Tolezano JE, Tolezano JE (1994) Ecoepidemiological aspects of american cutaneous leishmaniasis in the state of São Paulo, Brazil. *Memórias do Instituto Oswaldo Cruz*, **89**, 427–434.

Travi BL, Adler GH, Lozano M, Cadena H, Montoya-Lerma J (2002) Impact of habitat degradation on Phlebotominae (Diptera: Psychodidae) of tropical dry forests in Northern Colombia. *Journal of Medical Entomology*, **39**, 451–456.

Travi B, Jaramillo C, Segura S *et al.* (1994) *Didelphis Marsupialis*, an important reservoir of *Trypanosoma (Schizotrypanum) cruzi* and *Leishmania (Leishmania) chagasi* in Colombia. *The American journal of tropical medicine and hygiene*, 557.

Travi BL, Montoya J, Gallego J *et al.* (1996) Bionomics of *Lutzomyia evansi* (Diptera: Psychodidae) vector of visceral leishmaniasis in Northern Colombia. *Journal of Medical Entomology*, **33**, 278–285.

Uliana SR, Ruiz JC, Cruz AK (2008) *Leishmania* genomics: where do we stand? In: *Bioinformatics in Tropical Disease Research*, p. . National Center for Biotechnology Information, Bethesda (MD).

Van Buskirk J, Ostfeld RS (1995) Controlling Lyme disease by modifying the density and species composition of tick hosts. *Ecological Applications*, 1133–1140.

Vaz VC, D'Andrea PS, Jansen AM (2007) Effects of habitat fragmentation on wild mammal infection by *Trypanosoma cruzi*. *Parasitology*, **134**.

Volf P, Peckova J (2007) Sand flies and *Leishmania*: specific versus permissive vectors. *Trends in parasitology*, **23**, 91–92.

Volf P, Volfova V (2011) Establishment and maintenance of sand fly colonies. *Journal of Vector Ecology*, **36**, S1–S9.

Walsh JF, Molyneux DH, Birley MH (1993) Deforestation: effects on vector-borne disease. *Parasitology*, **106**, S55–S75.

Wasserberg G, Abramsky Z, Kotler BP *et al.* (2003) Anthropogenic disturbances enhance occurrence of cutaneous leishmaniasis in Israel deserts: patterns and mechanisms. *Ecological Applications*, **13**, 868–881.

Weiss RA (2001) The Leeuwenhoek lecture 2001. Animal origins of human infectious disease. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **356**, 957–977.

WHO (2016) Leishmaniasis in high-burden countries: an epidemiological update based on data reported in 2014. *Weekly epidemiological record*, **22**, 285–296.

WHO Expert Committee (2010) Control of the leishmaniases: report of a meeting of the WHO Expert Commitee on the Control of Leishmaniases, Geneva, 22-26 March 2010. *WHO Tech. Rep. Ser.*, **949**, 1–186.

Wolfe ND, Eitel MN, Gockowski J *et al.* (2000) Deforestation, hunting and the ecology of microbial emergence. *Global Change & Human Health*, **1**, 10–25.

Wood CL, Lafferty KD (2013) Biodiversity and disease: a synthesis of ecological perspectives on Lyme disease transmission. *Trends in Ecology & Evolution*, **28**, 239–247.

Wood CL, Lafferty KD, DeLeo G *et al.* (2014) Does biodiversity protect humans against infectious disease? *Ecology*, **95**, 817–832.

Wood CL, Lafferty KD, DeLeo G *et al.* (2016) Does biodiversity protect humans against infectious disease? Reply. *Ecology*, **97**, 543–546.

Woolhouse MEJ, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. *Emerging Infectious Diseases*, **11**, 1842–1847.

World Resources Institute (2014) Global Forest Watch.

Young DG, Duncan MA (1994) Guide to the identification and geographic distribution of *Lutzomyia* sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). *Memoirs of the American Entomological Institute*, **54**, 1–881.

Young H, Griffin RH, Wood CL, Nunn CL (2013) Does habitat disturbance increase infectious disease risk for primates? (K Lafferty, Ed,). *Ecology Letters*, **16**, 656–663.

Zelazny AM, Fedorko DP, Li L, Neva FA, Fischer SH (2005) Evaluation of 7sl RNA gene sequences for the identification of *Leishmania* spp. *The American Journal of Tropical Medicine and Hygiene*, **72**, 415–420.

Zink AR, Spigelman M, Schraut B *et al.* (2006) Leishmaniasis in Ancient Egypt and Upper Nubia. *Emerging Infectious Diseases*, **12**, 1616–1617.

Zinsstag J (2012) Convergence of Ecohealth and One Health. *Ecohealth*, **9**, 371–373.

CHAPITRE 1: CONTRIBUTION A LA CONNAISSANCE DE L'HISTOIRE ÉVOLUTIVE DES PHLÉBOTOMES: UTILISATION DE L'APPROCHE DE *GENOME SKIMMING*

RÉSUMÉ :

La bonne connaissance des liens de parenté existant entre les espèces constitue un atout considérable pour l'appréhension de leurs caractéristiques biologiques et écologiques. Malgré l'importance médicale des phlébotomes, l'histoire évolutive du groupe est mal connue. Pour les phlébotomes du nouveau monde, le seul cadre phylogénétique existant repose sur une analyse cladistique de caractères morphologiques datant d'il y a plus de 20 ans (Galati 1995), et qui n'a toujours pas été réellement confronté à des données moléculaires. Actuellement, la classification de Galati coexiste avec la classification phénétique de Young & Duncan (1994) qui est toujours largement employée, générant de la confusion dans la littérature.

Le *genome skimming* est une approche consistant à l'utilisation des technologies de séquençage haut-débit pour l'assemblage de régions génomiques répétées avec une faible profondeur de séquençage, et donc, pour un cout modéré, tout en s'affranchissant de longues étapes de PCR. Développée à l'origine pour le séquençage de génome d'organelles de plantes, la méthode a ensuite été appliquée pour le génome mitochondrial de différentes espèces animales.

Ce chapitre est constitué de trois articles. **Dans un premier article**, nous détaillons la méthodologie employée sur l'exemple du génome mitochondrial de la punaise réduve *Brontostoma colossus*, et proposons l'utilisation des données obtenues pour une analyse phylogénétique des hétéroptères. **Dans un deuxième article**, nous appliquons la même approche pour le séquençage du génome mitochondrial et des gènes des ribosomes nucléaires de *Nyssomyia umbratilis*, le principal vecteur de *Leishmania guyanensis*. Il s'agit du premier mitogénome de phlébotome séquencé. **Dans un troisième article**, le *genome skimming* est employé sur 13 espèces de phlébotomes néotropicaux supplémentaires. Le jeu de données génétiques obtenu est utilisé pour une analyse phylogénétique du groupe.

Malgré un jeu de données taxonomiquement restreint, les résultats de l'analyse sont statistiquement solides et semblent globalement supporter la classification cladistique de Galati qui sera employée préférentiellement dans les articles que nous avons écrits par la suite. De plus, l'emploi du *genome skimming* a permis de générer de précieuses données génétiques de référence qui pourront être employées lors de futurs développements méthodologiques. Ainsi, bien qu'il ne se rapporte pas directement à la question centrale de la thèse, ce premier chapitre apporte des éléments intéressants pour la suite de l'étude.

ARTICLE 1:

SHOTGUN ASSEMBLY OF THE ASSASSIN BUG *BRONTOSTOMA COLOSSUS* MITOCHONDRIAL GENOME (HETEROPTERA, REDUVIIDAE)

Arthur Kocher[1], Maria Kamilari[1,2], Emeline Lhuillier[3], Eric Coissac[4], Julie Péneau[5], Jérôme Chave[1], Jerôme Murienne[1*]

Published in *Gene* (doi: 10.1016/j.gene.2014.09.033)

*[1]Laboratoire EDB (UMR 5174, CNRS/UPS/ENFA) ; Université Paul Sabatier, Bât. 4R1, 118 route de Narbonne, 31062 Toulouse cedex 9, France ; [2]Division of Animal Biology, Department of Biology, School of Natural Sciences, University of Patras, GR-26500 Patras, Greece; [3]INRA, GeT–PlaGe, UAR 1209 Département de génétique animale, INRA Auzeville, 31326 Castanet-Tolosan, France ; [4]Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, 38041 Grenoble, France ; [5]Équipe « Écosystèmes Amazoniens et Pathologie Tropicale » (EA 3593), Faculté de Médecine Hyacinthe Bastaraud de l'Université des Antilles et de la Guyane, Campus Saint Denis, av. d'Estrées, F-97306 Cayenne, French Guiana.

ABSTRACT

The complete mitochondrial genome of the assassin bug *Brontostoma colossus* (Distant, 1902) (Heteroptera: Reduviidae) has been sequenced using a genome-skimming approach on an Illumina Hiseq 2000 platform. Fifty-four additional heteropteran mitogenomes, including five assassin bug species, were retrieved to allow for comparisons and phylogenetic analyses. The mitochondrial genome of *B. colossus* was determined to be 16,625 bp long, and consists of 13 protein-coding genes (PCGs), 23 transfer-RNA genes (tRNAs), two ribosomal-RNA genes (rRNAs), and one control region. The nucleotide composition is biased toward adenine and thymine (A+T=73.4%). Overall, architecture, nucleotide composition and genome asymmetry are similar among all available assassin bugs mitogenomes. All PCGs have usual start-codons (Met and Ile). Three T and two TA incomplete termination codons were identified adjacent to tRNAs, which was consistent with the punctuation model for primary transcripts processing followed by 3' polyadenylation of mature mRNA. All tRNAs exhibit the classic clover-leaf secondary structure except for $tRNA_{Ser(AGN)}$ in which the DHU arm forms a simple loop. Two notable features are present in the *B. colossus* mitogenome: *(i)* a 131 bp duplicated unit including the complete $tRNA_{Arg}$ gene, resulting in 23 potentially functional tRNAs in total, and (ii) an 857 bp duplicated region comprising 277 bp of the srRNA gene and 580 bp of the control region. A phylogenetic analysis based on 55 true bugs mitogenomes confirmed that *B. colossus* belongs to Reduviidae, but contradicted widely accepted hypothesis. This highlights the limits of phylogenetic analyses based on mitochondrial data only.

INTRODUCTION

Mitochondrial DNA has various interesting properties such as abundance in animal tissues, small size, relatively simple genomic structure, fast rate of evolution, and a straightforward mode of transmission with a low level of recombination (due to its maternal inheritance). This makes it a valuable tool for comparative genomics, population genetics and phylogenetics at various taxonomic resolutions (Avise et al., 1987; Moritz et al., 1987). An increasing number of complete mitochondrial genomes has been made available in the past decade, relying on long range Polymerase Chain Reaction (PCR), but this approach is difficult to perform and time-consuming. The immense yield now provided by Next Generation Sequencing (NGS) helps resolve these issues. The sequencing of the complete nuclear genome remains expensive because it requires a deep sequencing, but a relatively shallow sequencing can be used to recover the high copy fraction of mitochondrial DNA. This «genome skimming» approach, originally developed for plant organelles assemblage (Besnard et al., 2013; Malé et al., 2014; Straub et al., 2012), has been successfully used to assemble a wide variety of animal mitochondrial genomes (Besnard et al., 2014; Doyle et al., 2014; Thompson et al., 2014; Veale et al., 2014). However, mitogenome-based studies have been unbalanced among taxa, and the amount of available data for insects remains limited in comparison with that of vertebrates (Gissi et al., 2008; Salvato et al., 2008).

As for most metazoans, the mitogenome of insects is a circular double-stranded molecule of 14-20kb in size and exhibits a typical set of 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs) and 2 ribosomal RNA genes (rRNAs) (Boore, 1999; Wolstenholme, 1992), even though variations in gene content exist (Gissi et al., 2008; Junqueira et al., 2004; Shao and Barker, 2003; Thao et al., 2004). In addition, it contains a large non-coding region, the control region (CR), which is implicated in the initiation of transcription and replication processes (Bernt et al., 2013a; Clayton, 1992; Saito, 2005; Zhang and Hewitt, 1997).

Heteroptera (true bugs) contains over 40,000 described species to date, constituting one of the most diverse group of non-holometabolous insects (Weirauch and Schuh, 2011). Assassin bugs (Reduviidae) are a large family of mostly predatory land bugs belonging to the infraorder Cimicomorpha. It currently comprises close to 7,000 species worldwide, that exhibit a remarkable diversity in morphological traits and life habits (Weirauch and Munro, 2009; Wheeler, 1997). Some of them are of agricultural or medical importance, the most notorious being part of the hematophagous Triatominae subfamily, known as vectors of

Chagas disease in Central and South America (Lent and Wygodzinsky, 1979). Five assassin bugs mitogenomes have been sequenced so far: *Agriosphodrus dohrni* (Li et al., 2011), *Oncocephalus breviscutum* (Li et al., 2013), *Sirthenea flavipes* (Gao et. al, 2013), *Valentia hoffmanni* (Hua et. al, 2009) and *Triatoma dimidiata* (Dotson and Beard, 2001). Among Reduviidae, the Neotropical genus *Brontostoma* currently includes around 20 species (Dougherty, 1995; Gil-Santana and Baena, 2009; Maldonado Capriles, 1990). It is characterized by a bright coloration with red, yellow, black and brown. Like all members of the Ectrichodiinae subfamily, they are predators specialized on millipedes.

In this paper, we describe a genome-skimming approach using Illumina technology to assemble the complete mitochondrial genome of *Brontostoma colossus* (Distant, 1902). Its organization and features are compared to five other mitogenomes of Reduviidae. Fifty-four additional heteropteran mitochondrial genomes are used to perform a phylogenetic analysis.

MATERIAL AND METHOD

Specimen, DNA extraction and sequencing

One specimen of *B. colossus* was collected in French Guiana (RN2 Roura-Saint Georges) on April 20th 2010. Total genomic DNA was extracted from leg muscle tissue using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, USA), following a protocol adapted from the manufacturer's instructions (Appendix 1). The quality and quantity of extracted genomic DNA was evaluated using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and a PicoGreen double-stranded DNA quantitation assay kit (Life Technologies, Carlsbad, CA, USA).

The genomic DNA was sent for library construction and sequencing to the GeT-PlaGe core facilities of Genotoul (Toulouse, France). 288 ng of DNA were used for library construction using the Illumina TruSeq Nano DNA Sample Prep Kit following the instructions of the supplier (Illumina Inc., San Diego, CA, USA). After shearing by ultrasonication with a Covaris M220 (Covaris Inc., Woburn, MA, USA), purified fragments were A-tailed and ligated to sequencing indexed adapters. Fragments with an insert size around 450 bp were selected with Agencourt Ampure XP beads (Beckman Coulter, Inc.), and enriched with 8 cycles of PCR before library quantification and validation. The library was multiplexed with 23 other libraries (generated in other projects). The pool of libraries was then hybridized on one lane of Hiseq 2000 flow cell using the Illumina TruSeq PE Cluster Kit v.3, and paired-end reads of 100 nucleotides were collected on the HiSeq 2000 sequencer

using the Illumina TruSeq SBS Kit v.3 (200 cycles). Quality filtering was performed by the Consensus Assessment of Sequence and Variation (CASAVA) pipeline. Sequence data were stored on the NG6 platform (Mariette et al., 2012) and all the computations were performed on the computer cluster of the Genotoul bioinformatic platform (Toulouse, France).

Sequence assembly

Mitochondrial genome and nuclear ribosomal clusters were assembled using a previously described strategy (Besnard et al 2013; Besnard et al 2014; Malé et al. 2014). It is in essence similar to that proposed by Hahn et al. (2013). Reads aligning with the mitochondrial protein sequences of the closely related species *Triatoma dimidiata* (NC 002609, Dotson and Beard, 2001) were identified using the PLAST program (Nguyen and Lavenier, 2009). For the nuclear ribosomal cluster, we started from reads aligning with the 28S and 18S rRNA genes of *Eurydema maracandica* (Yu et al., 2013).

Reads with a match of at least 90% were assembled into contigs using the Velvet assembler (Zerbino and Birney, 2008) with a k-mer length of 81 and all the remaining parameters left at their default values. The resulting contigs were used as seeds to initiate the genome walking strategy (iterative mapping) using the extractreads2 program (included in the Obitools package; http://metabarcoding.org/obitools). Reads sharing at least 80 consecutive bp with the seeds were selected and subsequently used as seed to repeat the operation until no new read was identified. The newly selected reads were assembled with the Velvet assembler. The few resulting contigs were assembled using Geneious 6.0.6 Pro (Biomatters, Auckland, New Zealand). Two regions of the genome were not assembled using this procedure, due to ambiguities in the assembling process. At these locations, repeated regions were revealed by the coexistence of two assembling paths with significant coverage support (i.e. of the same magnitude than the average coverage), one path being at the junction between two copies of the repeated element, the other being at the end of the last copy (see Figure 1). We inferred the repeated copy number by comparing the number of reads mapping on a DNA fragment present only once in the mitogenome to that of DNA fragment belonging to the repeated element, assuming that the read coverage of a particular genomic region is proportional to its copy fraction in the sample. This coverage analysis is detailed in Appendix 2.

Coverage statistics were computed on the assembled genome with Geneious 6.0.6, by mapping the reads using the following mapping parameter: a minimum overlap of 100 bp, a minimum overlap identity of 95%, a word length of 50 and a maximum mismatch per read of 5%.

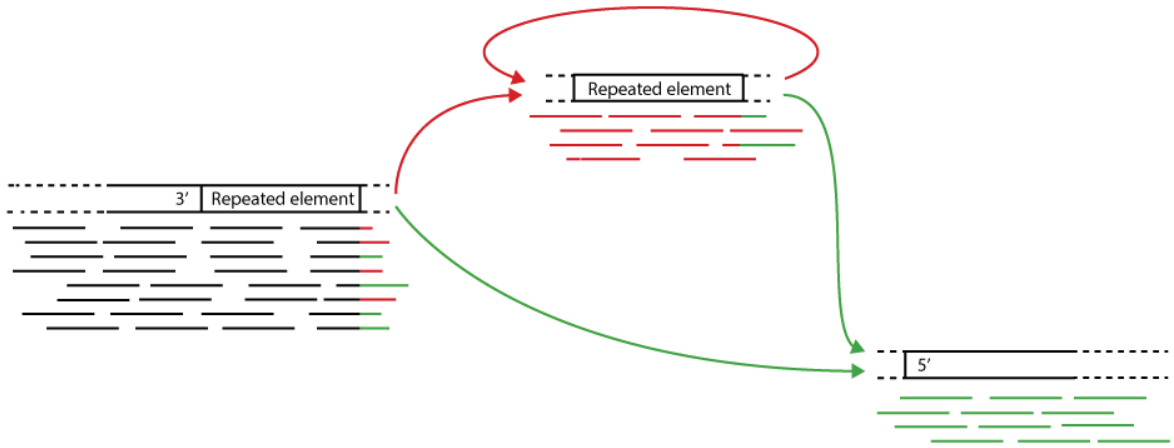


Figure 1: Schematization of an assembly ambiguity revealing a repeated element in the sequence. When assembling the 5' end of the repeated element, two different "assembly paths" are supported by a significant number of reads: (i) a path leading to the beginning of a novel repeated element (depicted in red) (ii) a path leading to the region flanking the 5' end of the last repetition (depicted in green).

Genome annotation

The mitochondrial genome was first annotated using the MITOS web server (Bernt et al., 2013b) applying the invertebrate mitochondrial genetic code (NCBI code Table 5). The annotations of tRNA genes were kept unchanged. The annotations of PCGs were refined by checking manually for consistent start/stop codons and reading frames. The annotations of rRNA genes were extended until adjacent tRNAs ($tRNA_{Ser(TCN)}$, $tRNA_{Leu(CUN)}$ and $tRNA_{Val}$), following the punctuation model of mtDNA transcription (Ojala et al., 1981; Stewart and Beckenbach, 2009). The 5' end of srRNA is not flanked by a tRNA. We adjusted it by mapping the srRNA of the *Agriosphodrus dohrni* mitogenome for which the secondary structured was predicted, and presented all expected domains (Li et al. 2011). We used Geneious 6.0.6 Pro with the following parameters: a word length of seven; maximum gap size of 15 and maximum mismatches of 40%. This approach conducted to extend lrRNA's annotation by 33 bp until $tRNA_{Leu}$ at its 3' end and by 578 bp until tRNA-Val at its 5' end. srRNA's annotation was extended by 2 bp until $tRNA_{Val}$ and by 47 bp at its 5' end. We further verify the consistency of these new annotations by mapping the rRNAs of *Agriosphodrus dohrni* as described above. The remaining large non-annotated sequence was annotated as the control region in homology with other insect mitogenomes.

The 18S and 28S rRNAs were annotated in comparison with that of *Eurydema maracandica* (Yu et al., 2013). The 5.8S rRNA was annotated in comparison with that of *Triatoma dimidiata* (accession number: KF142517).

Sequence analyses and phylogenetics

Base composition and codon usage where computed with MEGA6 (Tamura et al., 2013). AT-skew [(A-T)/(A+T)] and GC-skew [(G-C)/(G+C)] where used to measure nucleotide compositional differences between genes (Perna and Kocher, 1995) Relative synonymous codon usage (RSCU) where used to described bias in synonymous codon composition.

Tandem repeats were identified using Tandem Repeat Finder webserver (Benson, 1999). The secondary structure of tRNA's was inferred via the MITOS web server pipeline. Putative stem-loop structures were inferred using the RNAstructure web server (Bellaousov et al., 2013). We looked for structures conserved among the six assassin bugs within the 100 bp of the control region flanking $tRNA_{Ile}$ where stem loops structures have already been reported in Reduviidae (Dostson and Beard, 2001; Gao et al. 2013). We used the Turbofold algorithm (Harmanci et al. 2011), which infers secondary structure from high base pairing probabilities using the information derived from the sequence itself via the nearest neighbour thermodynamic model and also the information computed by using pairwise-sequence-alignment-based probabilities.

Fifty-four addditional heteropteran mitogenomes, including five assassin bugs, were downloaded from Genbank (Table 1). Two mitogenomes of Auchenorrhyncha were used as outgroups. The 13 PCGs were used for the analysis to allow for comparison with previous studies (Li et al., 2011; Yang et al., 2013). They were first aligned separately based on amino-acids translation with translatorX (Abascal et al., 2010). Divergent regions were removed with Gblocks.0.91b before back-translation in order to conserve reading frames. All resulting alignments were then concatenated using FASconCAT (Kück and Meusemann, 2010).The best partitioning scheme and substitution model were inferred with PartitionFinder.1.1.1 (Lanfear et al., 2012), using the greedy algorithm for scheme search and the Bayesian information criterion for scheme selection. A maximum-likelihood (ML) analysis was performed with RAxML 8.0 (Stamatakis, 2014), using the rapid bootstrap analysis option with the majority-rule tree based bootstopping criteria. Bootstrap support values were printed on the best ML tree. A Bayesian analysis was conducted using Mr.Bayes 3.2 (Ronquist and Huelsenbeck, 2003), starting from four random trees with 10 Markov chains (nine heated chain and 1 cold chain), 2,000,000 generations and all other parameters set to default. Each set was sampled every 200 generations with a burn-in of 25% of the sampled trees. At the end

Table 1: Complete or near-complete mitochondrial genomes used in this study

Suborder	Infra-order/ Superfamily	Family	Species	Accession num.	Reference
Auchenorryncha	Fulgoroidea	Fulgoridae	*Lycorma delicatula*	EU909203	(Song et al., 2012)
		Flatidae	*Geisha distinctissima*	NC_012617	(Song and Liang, 2009)
Heteroptera	**Cimicomorpha**				
	Cimicoidea	Anthocoridae	*Orius niger*	EU427341	(Hua et al., 2008)
	Miroidea	Miridae	*Adelphocoris fasciaticollis*	NC_023796	(Wang et al., 2014)
			Apolygus lucorum	NC_023083	(Wang et al., 2013)
			Lygus lineolaris	NC_021975	Unpublished
			Nesidiocoris tenuis	NC_022677	(Dai et al., 2012)
		Tingidae	*Corythucha ciliata*	NC_022922	(Yang et al., 2013)
	Naboidea	Nabidae	*Alloeorhynchus bakeri*	HM235722	(Li et al., 2012a)
			Gorpis annulatus	JF907591	(Li et al., 2012a)
			Gorpis humeralis	JF927830	(Li et al., 2012a)
			Himacerus apterus	JF927831	(Li et al., 2012a)
			Himacerus nodipes	JF927832	(Li et al., 2012a)
			Nabis apicalis	JF907590	(Li et al., 2012a)
	Reduvioidea	Reduviidae	*Agriosphodrus dohrni*	NC 015842	(Li et al., 2011)
			Brontostoma colossus	KM044501	This study
			Oncocephalus breviscutum	NC 022816	(Li et al., 2013)
			Sirthenea flavipes	HQ645959	(Gao et al., 2013)
			Triatoma dimidiata	NC 002609	(Dotson and Beard, 2001)
			Valentia hoffmanni	NC 012823	(Hua et al., 2009)
	Enicocephalomorpha				
	Enicocephaloidea	Enicocephalidae	*Stenopirates* sp.	NC_016017	(Li et al., 2012b)
	Gerromorpha				
	Gerroidea	Gerridae	*Aquarius paludum*	NC_012841	(Hua et al., 2009)
	Hydrometroidea	Hydrometridae	*Hydrometra* sp.	NC_012842	(Hua et al., 2009)
	Leptopodomorpha				
	Leptopodoidea	Leptopodidae	*Leptopus* sp.	FJ456946	(Hua et al., 2009)
	Saldoidae	Saldidae	*Saldula arsenjevi*	EU427345	(Hua et al., 2008)
	Nepomorpha				
	Corixoidea	Corixidae	*Sigara septemlineata*	FJ456941	(Hua et al., 2009)
	Naucoroidea	Aphelocheiridae	*Aphelocheirus ellipsoideus*	FJ456939	(Hua et al., 2009)
		Naucoridae	*Ilyocoris cimicoides*	NC_012845	(Hua et al., 2009)
	Nepoidea	Belostomatidae	*Diplonychus rusticus*	FJ456940	(Hua et al., 2009)
		Nepidae	*Laccotrephes robustus*	NC_012817	(Hua et al., 2009)
	Notonectoidea	Notonectidae	*Enithares tibialis*	NC_012819	(Hua et al., 2009)
	Ochteroidea	Gelastocoridae	*Nerthra* sp.	NC_012838	(Hua et al., 2009)
		Ochteridae	*Ochterus marginatus*	NC_012820	(Hua et al., 2009)
	Pleoidea	Helotrephidae	*Helotrephes* sp.	FJ456951	(Hua et al., 2009)
	Pentatomomorpha				
	Aradoidea	Aradidae	*Aradacanthia heissi*	HQ441233	(Shi et al., 2012)
			Brachyrhynchus hsiaoi	NC_022670	(Li et al., 2014)
			Neuroctenus parus	EU427340	(Hua et al., 2008)
	Coreoidea	Alydidae	*Riptortus pedestris*	EU427344	(Hua et al., 2008)
		Coreidae	*Hydaropsis longirostris*	EU427337	(Hua et al., 2008)
		Rhopalidae	*Aeschyntelus notatus*	EU427333	(Hua et al., 2008)
			Stictopleurus subviridis	NC_012888	(Hua et al., 2009)
	Lygaeoidea	Berytidae	*Yemmalysus parallelus*	EU427346	(Hua et al., 2008)
		Colobathristidae	*Phaenacantha marcida*	EU427342	(Hua et al., 2008)
		Geocoridae	*Geocoris pallidipennis*	EU427336	(Hua et al., 2008)
		Malcidae	*Chauliops fallax*	NC_020772	(Hua et al., 2008)
			Malcus inconspicuus	EU427339	(Hua et al., 2008)
	Pentatomoidea	Cydnidae	*Macroscytus gibbulus*	EU427338	(Hua et al., 2008)
		Dinidoridae	*Coridius chinensis*	JQ739179	(Liu et al., 2012)
		Pentatomidae	*Dolycoris baccarum*	NC_020373	(Zhang et al., 2013)
			Halyomorpha halys	NC_013272	(Lee et al., 2009)
			Nezara viridula	NC_011755	(Hua et al., 2008)
		Plataspidae	*Coptosoma bifaria*	EU427334	(Hua et al., 2008)
			Megacopta cribraria	NC_015842	(Hua et al., 2008)
		Tessaratomidae	*Eusthenes cupreus*	NC_022449	(Song et al., 2013)
		Urostylididae	*Urochela quadrinotata*	NC_020144	(Yuting et al., 2012)
	Pyrrhocoroidea	Largidae	*Physopelta gutta*	EU427343	(Hua et al., 2008)
		Pyrrhocoridae	*Dysdercus cingulatus*	EU427335	(Hua et al., 2008)

of the analysis, the average standard deviation of split frequencies was below the recommended 0.01.

Genome sequencing, assembly and annotation

After filtering 4.74% of the initial reads, raw sequence data represented a total of 7,831,929 paired-end reads (15,663,858 reads in total). Among the remaining reads, 34,224 were assembled into a 16,625 bp circular sequence, representing the complete mitochondrial genome with an average sequencing depth of 209.6. A circular map of the mitogenome and the assembly coverage are presented on Figure 2. The sequence was deposited in Genbank under the accession number KM044501. A total of 38 genes (13 PCGs, 23 tRNAs, two rRNAs) and one control region were identified. Twenty-four genes are encoded on the majority strand and the others mapped to the minority strand. Seven gene overlaps where observed, the longest being a 8-bp region between $tRNA_{Cys}$ and $tRNA_{Trp}$ (Table 2), which is a peculiar feature in Arthropoda (Bernt et al., 2013a). Apart from the control region, 13 non-coding regions ranging from 1 bp to 46 bp where identified.

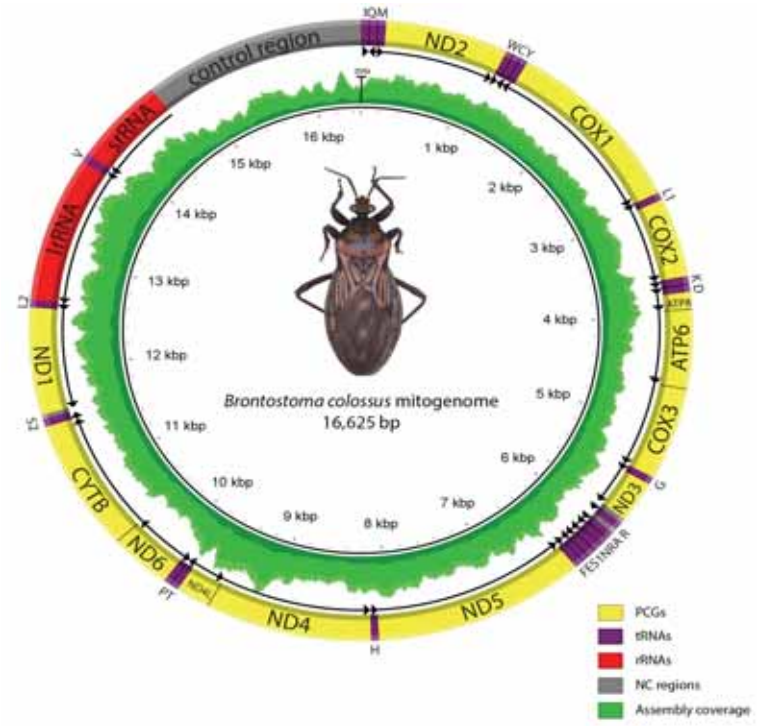


Figure 2: Schematic representation of Brontostoma colossus mitogenome. tRNAs are labelled according to the IUPAC-IUB single-letter amino acid codes. Arrows indicate directions of genes. The scale on the assembly coverage ring indicates 200x coverage.

Three repeated regions were identified: a 131-bp element containing $tRNA_{Arg}$; an 857-bp element consisting of 277 bp of srRNA and 580 bp of the control region and a 74-bp repeated element followed by a 39 partial copy located in the control region were tandem repeats are also found in other assassin bugs (Li et al., 2011). The existence of these duplications was supported by assembling paths with high coverage (over 100 reads). Details

on the results of the coverage analysis are given in Appendix 2. However, the sequencing technology used in this study does not allow inferring repeats copy numbers with high certainty, especially if a polymorphism exists in the sample. Indeed, heteroplasmy is often associated with tandem repetition (Zhang, 1997).

Table 2: Summary of the mitochondrial genome of *Brontostoma colossus*

Locus	Direction	Location (bp)	Size (bp)	Anticodon	Start codon	Stop codon	Interlocus nucleotides
$tRNA_{Ile}$	F	1-63	63	GAT			0
$tRNA_{Gln}$	R	66-134	69	TTG			2
$tRNA_{Met}$	F	133-202	70	CAT			-2
ND2	F	203-1193	991		ATT	T--	0
$tRNA_{Trp}$	F	1194-1260	67	TCA			0
$tRNA_{Cys}$	R	1253-1317	65	GCA			-8
$tRNA_{Tyr}$	R	1319-1382	64	GTA			1
COX1	F	1384-2917	1534		ATG	T--	1
$tRNA_{Leu(UUR)}$	F	2918-2982	65	TAA			0
COX2	F	2983-3661	679		ATT	T--	0
$tRNA_{Lys}$	F	3662-3733	72	CTT			0
$tRNA_{Asp}$	F	3734-3796	63	GTC			0
ATP8	F	3797-3955	159		ATT	TAA	0
ATP6	F	3949-4620	672		ATG	TAA	-7
COX3	F	4620-5408	789		ATG	TAA	-1
$tRNA_{Gly}$	F	5413-5477	65	TCC			4
ND3	F	5478-5831	354		ATA	TAA	0
$tRNA_{Arg}$	F	5873-5937	65	TCG			41
$tRNA_{Ala}$	F	5984-6044	61	TGC			46
$tRNA_{Arg}$	F	6049-6113	65	TCG			4
$tRNA_{Asn}$	F	6118-6187	70	GTT			4
$tRNA_{Ser(AGN)}$	F	6187-6255	69	GCT			-1
$tRNA_{Glu}$	F	6258-6319	62	TTC			2
$tRNA_{Phe}$	R	6318-6388	71	GAA			2
ND5	R	6389-8094	1706		ATT	TA-	0
$tRNA_{His}$	R	8095-8158	64	GTG			0
ND4	R	8159-9493	1335		ATG	TAA	0
ND4L	R	9487-9771	285		ATT	TAA	-7
$tRNA_{Thr}$	F	9774-9837	64	TGT			2
$tRNA_{Pro}$	R	9838-9903	66	TGG			0
ND6	F	9906-10397	492		ATA	TAA	2
CYTB	F	10397-11529	1133		ATG	TA-	-1
$tRNA_{Ser(TCN)}$	F	11530-11598	69	TGA			0
ND1	R	11637-12548	912		ATT	TAA	38
$tRNA_{Leu(CUN)}$	R	12549-12613	65	TAG			0
lrRNA	R	12614-13864	1251				0
$tRNA_{Val}$	R	13865-13935	71	TAC			0
srRNA	R	13936-14728	793				0
Control region		14729-16625	1896				0

The complete nuclear ribosomal gene cluster was recovered. A total of 21,514 reads were assembled into an 8,287 bp sequence comprising of 18S rRNA (1,893 bp), ITS1 (1,141 bp), 5.8S rRNA (155 bp), ITS2 (941 bp) and 28S rRNA (4,077 bp). The sequence was deposited on Genbank under the accession number KM278219.

The mitogenome of *B. colossus* shares the same architecture and orientation as the other six mitogenomes of assassin bugs, except for the presence of an additional $tRNA_{Arg}$ gene, as will be described below. This gene arrangement (without the additional $tRNA_{Arg}$) is also found in *Drosophila melanogaster* and was the first to be determined, differing by a single tRNA translocation, from that of the chelicerate *Limulus polyphemus,* which is considered ancestral for Arthropoda (Boore et al., 1995; Lavrov et al., 2000). This mitogenome organisation is also found in crustaceans, and is thought to be ancestral for the insect-crustacean clade (Boore, 1999; Boore et al., 1998). Among the available data for heteropterans, seven out of 54 species have been found to present a different gene arrangement: *Nabicalis apicalis* mitogenome miss the cluster containing $tRNA_{Ile}$, $tRNA_{Gln}$ and $tRNA_{Met}$ (Li et al., 2012a). $tRNA_{Ile}$ and $tRNA_{Gln}$ are also missing in *Urochela quadrinotata* (Yuting et al., 2012). The positions of $tRNA_{Thr}$ and $tRNA_{Pro}$ are inverted in *Physopelta gutta* (Hua et al., 2008). The gene order of the *Stenopirates* sp. mitogenome differs largely with the inversion of two tRNA genes ($tRNA_{Thr}$ and $tRNA_{Pro}$) and translocations of five gene clusters ($tRNA_{Thr}$-$tRNA_{Pro}$ -ND6, CYTB-$tRNA_{Ser(TCN)}$, ND1-$tRNA_{Leu(CUN)}$, l-rRNA- $tRNA_{Val}$ -s-rRNA and the control region) between ND4L and $tRNA_{Ile}$ (Li et al., 2012b). The positions of $tRNA_{Cys}$ and $tRNA_{Trp}$ are exchanged in *Aradacanthia heissi* (Shi et al., 2012). In the latter, as well as in the two other Aradoidea represented (*Brachyrhynchus hsiaoi* and *Neuroctenus parus*); the positions of $tRNA_{Ile}$ and $tRNA_{Gln}$ are exchanged (Li et al., 2014; Hua et al., 2008).

The size of the six assassin bug mitogenomes ranges from 15,625 bp in *V. hoffmannii* to 17,019 bp in *T. dimidiata*. These differences are mostly due to variations in the size of the control region, which is generally observed for all insects. Previous studies have reported control region size ranging from 70 bp in *Ruspolia dubia* (Orthoptera) to 4,599bp in *Drosophila melanogaster* (Diptera) (Garesse, 1988; Zhou et al., 2007).

Protein-coding genes

The total length of the 13 PCGs was 11,041 bp. Their nucleotide composition is strongly biased toward AT with an overall AT content of 72.8% (Appendix 3). All PCGs have an ATN start codon (Table 2). Six PCGs initiated with ATT (ND2, COX2, ATP8, ND5, ND4L and ND1), five initiated with ATG (COX1, ATP6, COX3, ND4 and CYTB) and two initiated with ATA (ND3 and ND6). Four genes share the same ATG start codon in the six assassin bugs mitogenomes (COXI, ATP6, COXIII and ND4). No GTG start codon was found in *B. colossus* in contrast with other assassin bugs for ND5, ND4L and ND1 genes (Appendix 4). Other unconventional start codons were described in insects such as TTG in heteropterans (Yang et al., 2013), CGA and TTAG in lepidopterans (Lee et al., 2006; Yukuhiro et al., 2002), or ATAA, GTAA and TTAA in dipterans (Ballard, 2000; Clary and Wolstenholme, 1985) but none of them were found in assassin bugs.

The majority of PCGs have a usual TAA stop codon, but three T and two TA stop codons were identified (ND2, COX1, COX2 and ND5, CYTB respectively). These incomplete stop codons are immediately adjacent to tRNA genes encoded on the same strand, consistent with the punctuation model for primary transcripts processing followed by 3' polyadenylation of mature mRNA that will allow the completion of termination codons (Nagaike, 2005; Ojala et al., 1981; Stewart and Beckenbach, 2009). Incomplete stop codons can be found in all six assassin bugs mitogenomes and are shared with many arthropods (Boore, 2000).

Ribosomal and transfer RNA genes

rRNA genes locations and lengths are similar to those of other insects mitogenomes. lrRNA is located between $tRNA_{Leu(CUN)}$ and $tRNA_{Val}$ and is 1,251 bp-long. srRNA is located between $tRNA_{Val}$ and the control region and is 793 bp-long. Their AT content is respectively 79.0% and 74.0%.

The classical set of 22 tRNAs found in arthropods is present in *B. colossus,* but an additional copy of the $tRNA_{Arg}$ gene was identified (see below). Their lengths vary between 61 bp ($tRNA_{Ala}$) and 72 bp ($tRNA_{Lys}$). Secondary structures of tRNAs are schematized in Appendix 5. The classical clover leaf structure was observed for each of them, except for $tRNA_{Ser(AGN)}$, in which the D arm is reduced to a simple loop, as in many insects, and more generally, in most bilaterians (Bernt et al., 2013a; Wolstenholme, 1992).

Flanking $tRNA_{Ala}$, we identified a duplicated element consisting of two identical 131-bp copies separated by eight non-coding base pairs. It includes the entire $tRNA_{Arg}$ and short-flanking regions corresponding to 28 bp of $tRNA_{Ala}$ and 38 bp of $tRNA_{Asn}$ (Figure 3). This unusual feature results in two copies of $tRNA_{Arg}$ for a total of 23 tRNA genes which is relatively rare in insect mitogenomes, even though more than 22 tRNAs were already observed in other species such as *Coreana raphaelis* (Lepidoptera; Kim et al., 2006)), *Thrips imaginis* (Thysanoptera; Shao and Barker, 2003), *Chrysomya chloropyga* (Diptera; Junqueira et al., 2004) and *Trialeurodes vaporariorum* (Hemiptera; Thao et al., 2004). To our knowledge, the presence of an additional $tRNA_{Arg}$ has only been described in Porifera and Placozoa (Lavrov and Lang, 2005; Signorovitch et al., 2007). The two copies of this duplicated element are strictly identical, which suggests a recent origin of the duplication event. Interestingly, it could lead to gene rearrangement through a duplication/deletion mechanism involving the random deletion of the original copy of the gene (Boore, 2000; Moritz and Brown, 1986; Shao et al., 2006). Most repeated elements are located close to the replication origin, supporting the idea that mitogenomic duplication events are mainly due to replication slippage mechanisms (Macey et al., 1998; Zhang and Hewitt, 1997). However, to our knowledge, no putative replication origin located close to this duplicated region has been mentioned so far.

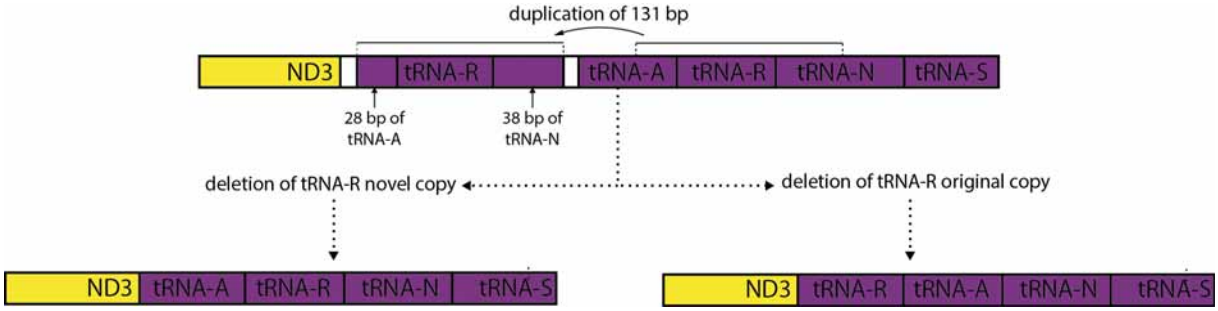


Figure 3: Organization of the 131 bp duplicated sequence comprising tRNA-Arg and potential gene rearrangements after random deletion of one tRNA-arg copy.

Non-coding regions

Thirteen short intergenic spacers (IGS) and a long control region were identified, matching the usual organisation of insect mitogenomes. Most IGS are very short, with less than four base pairs, and seven overlapping sequences are found (Table 2). The longest IGS are the 41 and 46 bp-flanking regions of the first copy of $tRNA_{Arg}$ and the 38-bp IGS found between ND1 and $tRNA_{Ser2}$. The latter is remarkably long in every assassin bug mitogenome, ranging from 23 bp in *V. hoffmanni* to 309 bp in *T. dimidiata*. It exhibits tandem repeats in *A.*

dohrni and *T. dimidiata* and has been suspected to be one of the replication origins (Dotson and Beard, 2001; Li et al., 2011) comparably to the 193 bp region located between the $tRNA_{leu(UUR)}$ and COXII genes in the honeybee *A. mellifera* (Crozier and Crozier, 1993).

The control region is 1,896-bp long and is located between the srRNA and $tRNA_{Ile}$ genes. In *B. colossus*, as well as in the five other Reduviidae, it exhibits a higher G+C content than that of the whole mitogenome, in contrast with other insect species in which the control region was found to be remarkably A+T rich (Zhang and Hewitt, 1997).

The alignment of the six Reduviidae control regions reveals a conserved sequence block (CSB) of 40 bp, including a string of 13 Gs (Figure 4a; Li et al., 2011). CSBs have been identified in the control region of various metazoans and are generally thought to play a role in the replication mechanism (Lee et al., 1995; Walberg and Clayton, 1981; Zhang and Hewitt, 1997). However, we did not find similarity between the CSB described here and those reported for other taxa, even though G islands have already been described in other insects (Oliveira et al., 2008). More studies are needed to identify precisely replication origins and to speculate on the role of adjacent sequences, as was proposed by (Saito, 2005) who strongly suspected a "T-strech" sequence conserved in *Drosophila* and other insect species to be involved in the replication process.

The control regions of the six assassin bugs present a similar organization (Figure 4b). In all of them, except for *S. flavipes,* tandem-repeats were identified between the CSB and $tRNA_{Ile}$. In *B. colossus,* they consist of four 74 bp units and one 39 bp unit, the latter corresponding to a partial copy of the 74 bp unit. The 100 bp preceding the CSB are remarkably G+C rich (42%) in *B. colossus*, but the nucleotide composition of the whole control region does not show any clear pattern, in contrast with other assassin bugs in which long G+C and A+T rich regions were found preceding and following the CSB respectively (Dotson and Beard, 2001; Gao et al., 2013; Li et al., 2011). In *B. colossus* control region, a notable feature is a large duplicated region consisting of two identical copies of 857 bp separated by seven non-coding nucleotides. It includes 277 bp of srRNA and 580 bp of the control region, comprising the CSB. It is remarkably long and results in two copies of the CSB, which could therefore have implications on the mitogenome replication process.

Within the 60 bp of the control region flanking the $tRNA_{Ile}$, DNA segments have the potential to form stem-loop structures involving at least 11 base-pairings in the six assassin bugs mitogenomes except for *V. hoffmanni* (Figure 5). Such features may be involved in the replication mechanism (Song and Liang, 2009; Zhang et al., 1995). However, this is only

speculative. More comprehensive studies would be required to assess the significance of these inferences and the potential role of these structures.

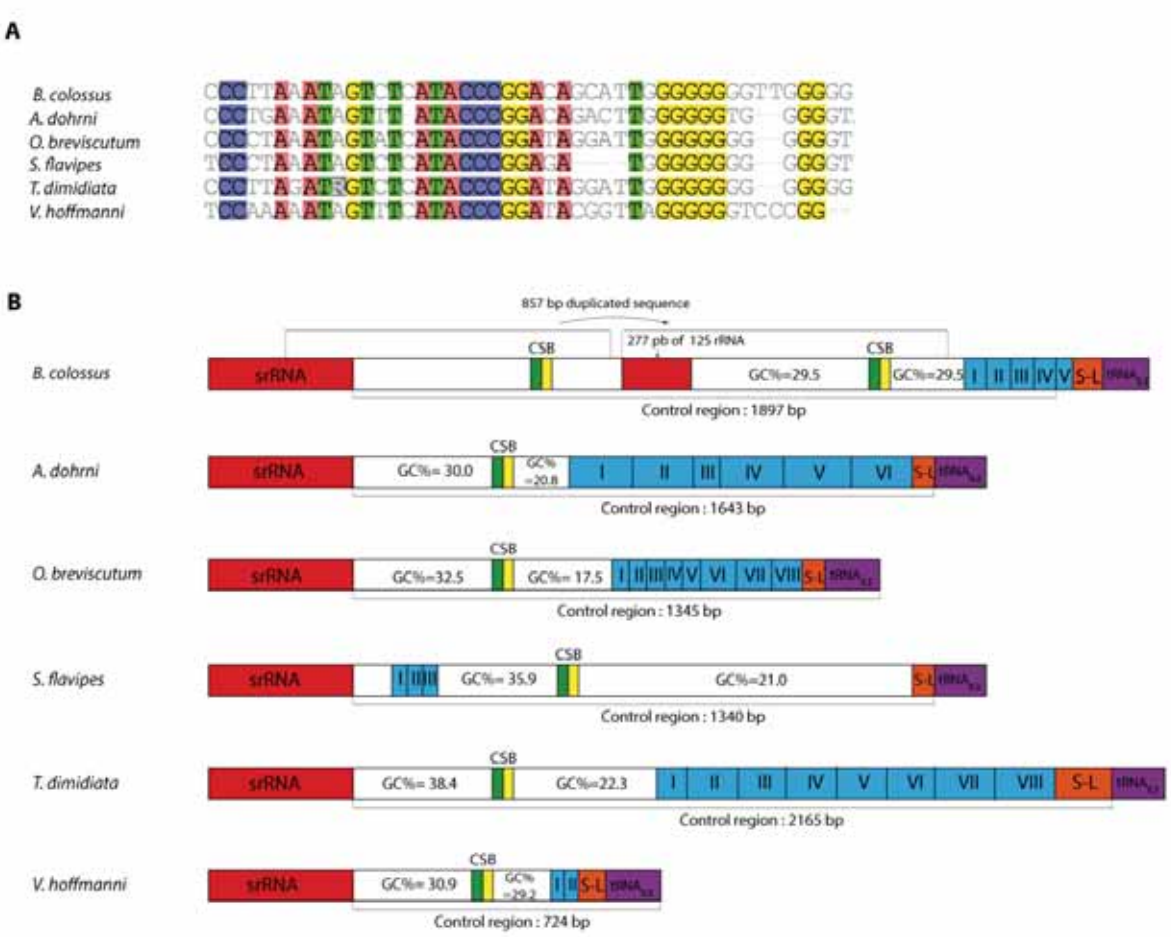


Figure 4: (A) Alignment of the conserved sequence blocks identified in the mitochondrial control region of the six assassin bugs. Nucleotide positions that are conserved among the six assassin bugs are highlighted. **(B)** Structural organization of the mitochondrial control region of the six assassin bugs. The blue boxes with roman numerals indicate tandem repeat units. The CSB box indicates the conserved sequence block, and the yellow part indicates the "G element". The orange box "S-L" indicates the region where potential stem loops are found. GC content is shown in the white boxes previously described as GC-rich and AT-rich regions.

Nucleotide content and codon usage

The nucleotide composition is strongly biased toward adenine and thymine in the mitogenome of *B. colossus*, with A+T representing 73.5% of the whole sequence and ranging from 70.2% in the control region, 70.8% in protein-coding genes, 76.8% in tRNA genes to 77.1% in rRNA genes. AT-rich codons are predominant, with the most prevalent being in order ATT (Ile), TTA (Leu), TTT (Phe) and ATA (Met). The relative synonymous codon usage (RSCU) clearly indicates that AT rich codons are favoured among synonymous codons Appendix 6. At the third codon position, AT content is particularly high (82.8%), and G

nucleotides are under-represented (GC skew= -0.20). AT content, as well as A-T and G-C skews patterns, are similar among the six assassin bugs mitochondrial genomes (Appendix 3).

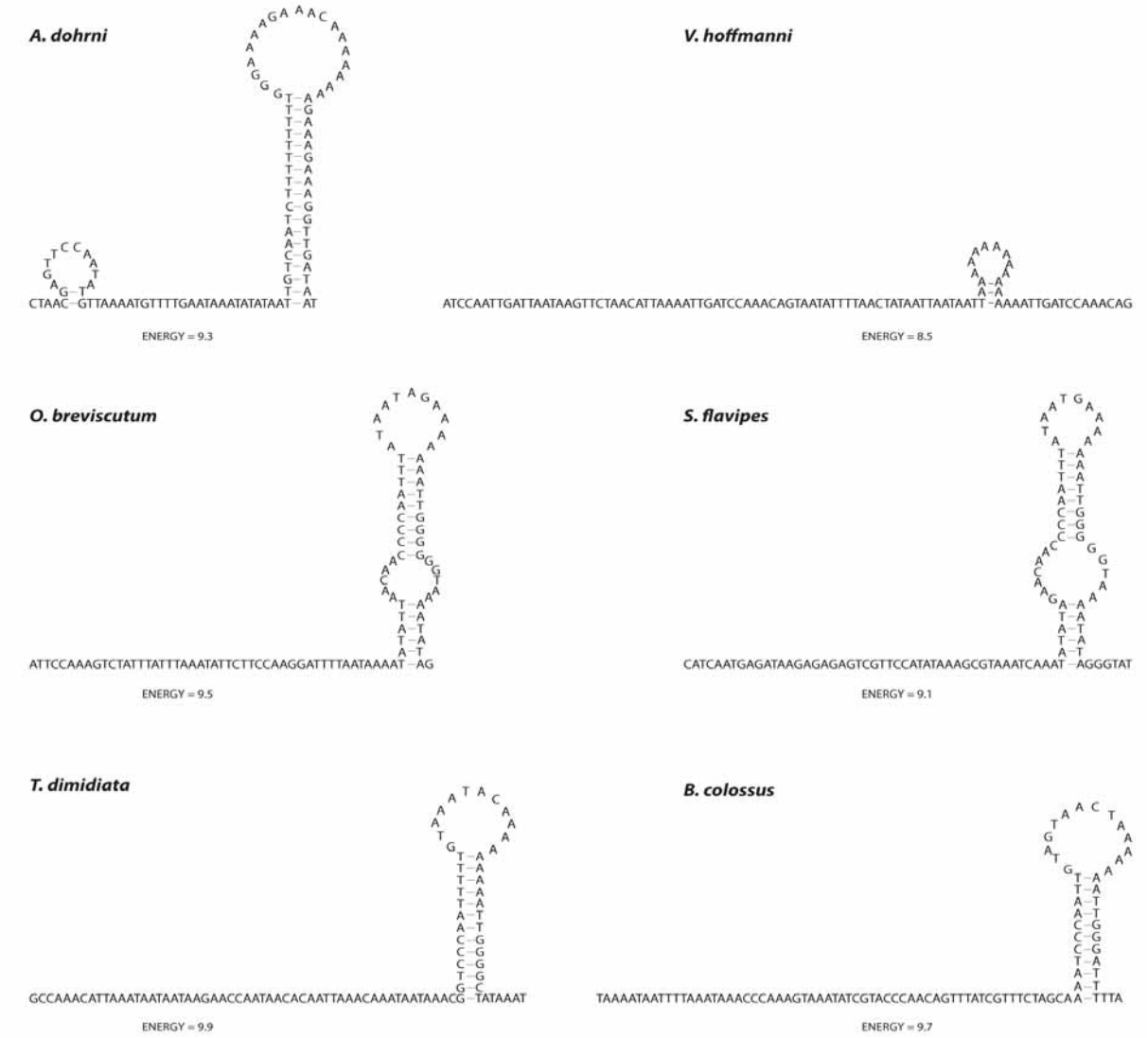


Figure 5: Conserved stem loop structures inferred by the Turbofold algorithm on the the 100 bp of the control region flanking $tRNA_{Ile}$ of the six assassin bugs. Values of free energy are indicated above each structure.

Phylogenetic analysis

Bayesian inference and Maximum Likelihood analysis (ML) generated phylogenetic trees with very similar topologies. The tree inferred by the Bayesian method is presented in Figure 6 with nodes posterior probabilities and ML bootstrap support values. The topology of the best ML tree for Reduviidae is also presented. The relationships among Reduviidae are conserved in both analyses except for the position of *B. colossus*, which is placed as a sister group of *V. hoffmanni* under Bayesian inference whereas it is the early lineage of Reduviidae in the ML tree. Our results are not well supported and are hardly comparable with those of recent studies that have addressed the relationships among assassin bugs based on nuclear and

mitochondrial DNA as well as morphological data for a large number of taxa (Weirauch and Munro, 2009; Weirauch, 2008). However, the higher-level relationships of Reduviidea remain poorly resolved and the addition of mitogenomic data for more taxa will surely provide useful phylogenetic information in the future.

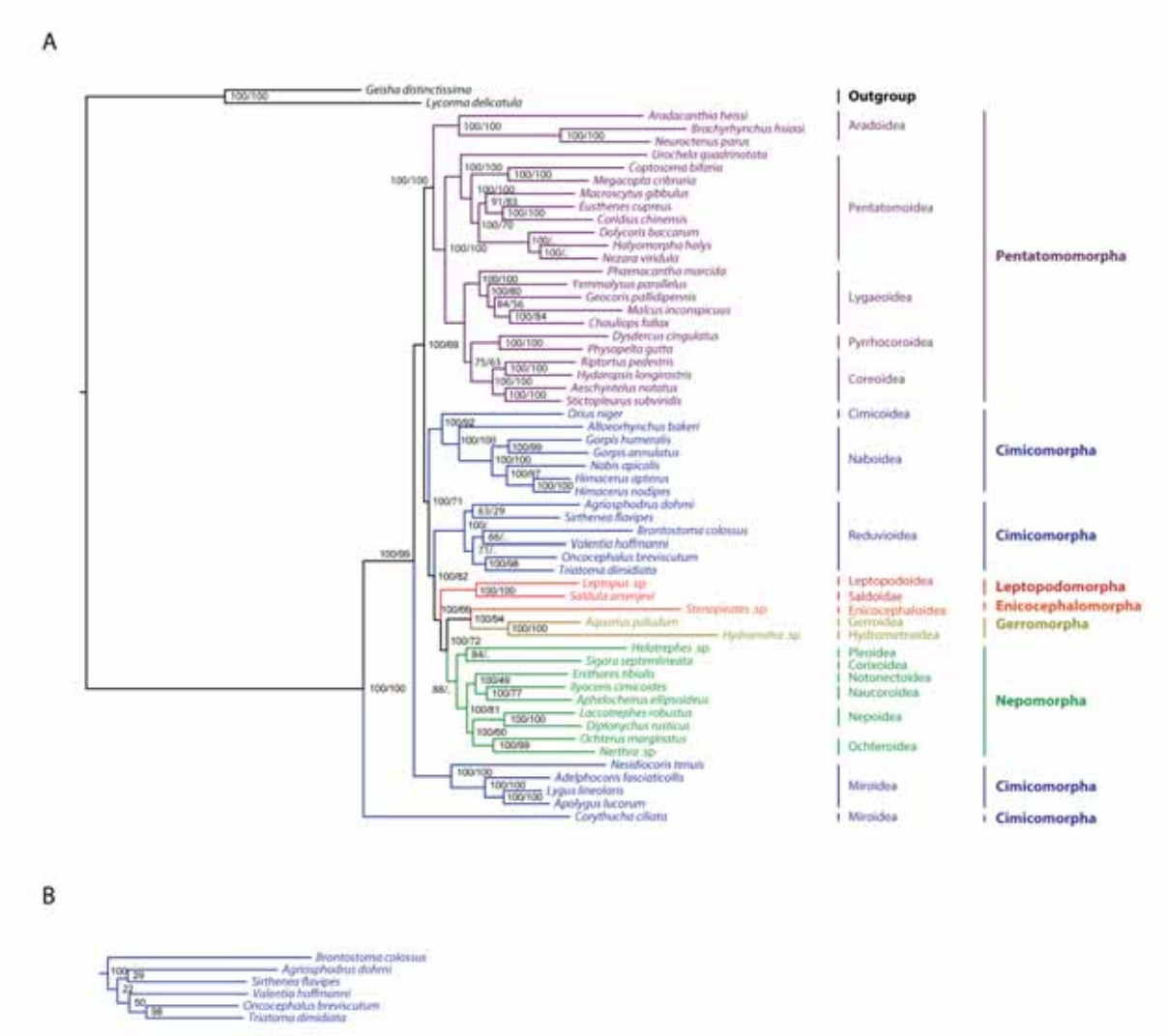


Figure 6: (A) Phylogenetic tree inferred by Bayesian analysis from 55 heteropteran mitogenomes. Left numbers at the nodes indicate Bayesian posterior probabilities expressed in percentages. When the node was also present on the tree inferred by ML analysis, right numbers indicate bootstrap support values. A dot indicates that the node was absent on the tree inferred by ML analysis **(B)** Maximum likelihood (ML) sub-tree corresponding to Reduviidae family, with bootstrap values depicted on nodes.

The 20 superfamilies represented in our dataset are monophyletic except for Miroidea. At the infra-order level, Pentatomorpha is monophyletic with the following relationships between the five superfamilies: Aradoidea + (Pentatomoidea + (Lygaeoidea + (Pyrrhocoroidea + Coreoidea))). These relationships are strongly supported in our analyses and do not confirm the results of previous studies based on a subset of the present mitogenomic data that placed Coreoidea and Lygaeoidea as sister groups (Hua et al., 2008;

Yang et al., 2013). Based on the tree generated by Bayesian inference, Nepomorpha is monophyletic with the following relationships between the six superfamilies: (Pleoidea + Corixoidea) + ((Notonectoidea + Naucoroidea) + (Nepoidea + Ochteroidea)). ML analysis differs in positioning Pleoidea as a sister group of the remaining Nepomorpha. These results are inconsistent with those of a recent study based on molecular and morphological data (Hebsgaard et al., 2004). Interestingly, they also contradict a previous analysis of a subset of the present mitogenomic dataset by confirming the monophyly of Nepomorpha including Pleoidea, for which an infraordinal status was proposed (Hua et al., 2009). Gerromorpha and Leptopodomorpha are also monophyletic, but only two species of each were included in the study. Enicocephalomoprha was only represented by *Stenopirates* sp.

Our analysis supports the paraphyly of Cimicomorpha that consisted of four different clades: (Cimicoidea + Naboidea), Reduvioidea, Miridae and Tingidae. Tingidae is placed as a sister group to all remaining Heteroptera. However, the monophyly of Cimicomorpha has been widely accepted and is supported by various analyses that have been using mitochondrial and nuclear data for a larger number of taxa (Li et al., 2012; Schuh et al., (2009); Tian et al., 2008). Infraordinal relationships are conserved in both ML and Bayesian analyses: Tingidae + (Miridae + (Pentatomomorpha + ((Cimicoidea + Naboidea) + (Leptopodomorpha + ((Enicocephalimorpha + Gerromorpha) + Nepomorpha))))). These results are questionning the general consensus that considers Enicocephalomorpha as the early infraorder of Heteroptera (Weirauch et al., 2011). However, the relationships between the infraorders of Heteroptera remain controversial. Only few phylogenetic studies have addressed the question and most of these have only included a small number of taxa (Mahner, 1993; Wheeler et al., 1993; Xie et al., 2008). Mitogenomic data provide a new insight in this regard, but more taxa should be added to the current database, especially in the poorly represented infra-orders Enicocephalomorpha, Gerromorpha, Leptopodomorpha and Dipsocoromorpha.

Our results are incongruent with current phylogenetic hypothesis of Heteroptera (e.g. Schuh et al. 2009). On the other hand, they are in accordance with previous studies based on a subset of the present mitogenomic data (Li et al., 2011; Yang et al., 2013). Analyses performed on individual genes by Tian et al (2008) and Schuh et al. (2009) indicate that the monophyly of Cimicomorpha is only supported by nuclear DNA. The incongruence of phylogenetic analyses among different genomic regions is a well-known issue that can have various biological causes such as incomplete lineage sorting or rate variation among partitions (Som, 2014). Lin and Danforth (2004) studied the differences in the pattern of nucleotide

substitution among nuclear and mitochondrial genes and concluded that insect phylogenetic studies should increasingly focus on nuclear data. This raises the limitations of phylogenetic inferences from complete mitochondrial genomes only. While reducing stochastic errors by providing large molecular datasets, this approach is susceptible to potential site-specific bias and would benefit from a conjoint analysis with other genes. One advantage of the genome skimming approach used in this paper is that it allows the recovery of nuclear genes of phylogenetic interest in addition to the full mitogenome sequence. However, 18S and 28S data available to date on Genbank are too scarce (14 species out of 55 represented in our mitochondrial dataset) to perform a combined analysis.

CONCLUDING REMARKS

This study provides further evidence that NGS can be used efficiently to generate mitogenomic data with a low amount of DNA. We successfully recovered the full mitogenome sequence of *Brontostoma colossus,* which included two unusual duplicated regions. The Illumina technology used in this study identified repeated elements without ambiguities, but their copy number can only be estimated using the sequence coverage information. In a near future, the rapid evolution of sequencing technologies (especially read length) and bioinformatics tools will probably bring improvements in this regard. The increasing number of full mitochondrial genome sequences brings precious phylogenetic information. However, our study highlights the limits of analyses based on mitochondrial DNA only. Currently, there is a lack of correspondence between publicly available mitogenomic and nuclear data. The genome skimming-approach could provide an interesting improvement in this regard. In a single experimentation, it allows to recover the full mitogenome sequence as well as nuclear genes that are classically used for phylogenetic inference. We argue that future studies reporting full organelles sequencing with genome skimming approaches should systematically report the assembly of the nuclear genes of phylogenetic interest.

ACKNOWLEDGEMENTS

This work was supported by PO-FEDER TIMGED N°30195 (Trypanosomes d'Intérêt Médical en Guyane française – Epidémiologie et Diagnostic) and "Investissement d'Avenir" grants managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01; TULIP, ANR-10-LABX-41, ANR-11-IDEX-0002-02) as well as project METABAR (ANR-

11-BSV7-0020). We would like to thank Christine Aznar, Denis Blanchet, Jean-Pierre Dujardin and Jean-Michel Bérenger for the help with obtaining and identifying the specimen. We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing computing and storage resources.

REFERENCES

Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38, W7–13.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18, 489–522.

Ballard, J.W.O., 2000. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. J. Mol. Evol. 51, 48–63.

Bellaousov, S., Reuter, J.S., Seetin, M.G., Mathews, D.H., 2013. RNAstructure: web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res. 41, W471–W474.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573.

Bernt, M., Braband, A., Schierwater, B., Stadler, P.F., 2013a. Genetic aspects of mitochondrial genome evolution. Mol. Phylogenet. Evol. 69, 328–338.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013b. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319.

Besnard, G., Christin, P.-A., Malé, P.-J.G., Coissac, E., Ralimanana, H., Vorontsova, M.S., 2013. Phylogenomics and taxonomy of Lecomtelleae (Poaceae), an isolated panicoid lineage from Madagascar. Ann. Bot. 112, 1057–1066.

Besnard, G., Jühling, F., Chapuis, É., Zedane, L., Lhuillier, É., Mateille, T., Bellafiore, S., 2014. Fast assembly of the mitochondrial genome of a plant parasitic nematode (*Meloidogyne graminicola*) using next generation sequencing. C. R. Biol. 337, 295–301.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27, 1767–1780.

Boore, J.L., 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In: Sankoff, D. and Nadeau, J.H. eds. Comparative Genomics. Springer Netherlands, pp. 133–147.

Boore, J.L., Collins, T.M., Stanton, D., Daehler, L.L., Brown, W.M., 1995. Deducing the pattern of arthropod phylogeny from mitochondrial-DNA rearrangements. Nature 376, 163-165.

Boore, J.L., Lavrov, D.V., Brown, W.M., 1998. Gene translocation links insects and crustaceans. Nature 392, 667–668.

Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22, 252–271.

Clayton, D.A., 1992. Transcription and replication of animal mitochondrial DNAs. Int Rev Cytol 141, 217–232.

Crozier, R.H., Crozier, Y.C., 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics 133, 97–117.

Dai, X., Xun, H., Chang, J., Zhang, J., Hu, B., Li, H., Yuan, X., Cai, W., 2012. The complete mitochondrial genome of the plant bug *Nesidiocoris tenuis* (Reuter) (Hemiptera: Miridae: Bryocorinae: Dicyphini). Zootaxa (3554), 30–44.

De Bruijn, M.H., 1983. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. Nature 304, 234-241.

Dotson, E.M., Beard, C.B., 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. Insect Mol. Biol. 10, 205–215.

Dougherty, V., 1995. A review of the new world Ectrichodiinae genera (Hemiptera: Reduviidae). Trans. Am. Entomol. Soc. 121,173–225.

Doyle, S.R., Griffith, I.S., Murphy, N.P., Strugnell, J.M., 2014. Low-coverage MiSeq next generation sequencing reveals the mitochondrial genome of the Eastern Rock Lobster, *Sagmariasus verreauxi*. Mitochondrial DNA 0,1–2.

Gao, J., Li, H., Truong, X.L., Dai, X., Chang, J., Cai, W., 2013. Complete nucleotide sequence and organization of the mitochondrial genome of *Sirthenea flavipes* (Hemiptera: Reduviidae: Peiratinae) and comparison with other assassin bugs. Zootaxa 3669, 1-16.

Garesse, R., 1988. *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations. Genetics 118, 649–663.

Gil-Santana, H., Baena, M., 2009. Two new species of *Brontostoma* Kirkaldy (Hemiptera: Heteroptera: Reduviidae: Ectrichodiinae) from Bolivia, with description of the male genitalia of two other species of the genus, and description of the female of *B. doughertyae* Gil-Santana, Lopes, Marques & Jurberg. Zootaxa 1979, 41–52.

Gissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101, 301–320.

Hahn C, Bachmann L, Chevreux B, 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research, 41, e129.

Harmanci, A.O., Sharma, G., Mathews, D.H., 2011. TurboFold: Iterative probabilistic estimation of secondary structures for multiple RNA sequences. BMC Bioinformatics 12, 108.

Hebsgaard, M.B., Andersen, N.M., Damgaard, J., 2004. Phylogeny of the true water bugs (Nepomorpha: Hemiptera–Heteroptera) based on 16S and 28S rDNA and morphology. Syst. Entomol. 29, 488–508.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., Bu, W., 2008. Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). BMC Genomics 9, 610.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., Bu, W., 2009. Phylogenetic analysis of the true water bugs (Insecta: Hemiptera: Heteroptera: Nepomorpha): evidence from mitochondrial genomes. BMC Evol. Biol. 9, 134.

Junqueira, A.C.M., Lessinger, A.C., Torres, T.T., da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo Espin, A.M.L., 2004. The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). Gene 339, 7–15.

Kim, I., Lee, E.M., Seol, K.Y., Yun, E.Y., Lee, Y.B., Hwang, J.S., Jin, B.R., 2006. The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). Insect Mol. Biol. 15, 217–225.

Kopp, A., True, J.R., 2002. Phylogeny of the Oriental *Drosophila melanogaster* species group: a multilocus reconstruction. Syst. Biol. 51, 786–805.

Kück, P., Meusemann, K., 2010. FASconCAT: Convenient handling of data matrices. Mol. Phylogenet. Evol. 56, 1115–1118.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.

Lavrov, D.V., Boore, J.L., Brown, W.M., 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Mol. Biol. Evol. 17, 813–824.

Lavrov, D.V., Lang, B.F., 2005. Transfer RNA gene recruitment in mitochondrial DNA. Trends Genet. 21, 129–133.

Lee, W.-J., Conroy, J., Howell, W.H., Kocher, T.D., 1995. Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. 41, 54–66.

Lee, E.-S., Shin, K.S., Kim, M.-S., Park, H., Cho, S., Kim, C.-B., 2006. The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). Gene 373, 52–57.

Lee, W., Kang, J., Jung, C., Hoelmer, K., Lee, S.H., Lee, S., 2009. Complete mitochondrial genome of brown marmorated stink bug *Halyomorpha halys* (Hemiptera: Pentatomidae), and phylogenetic relationships of hemipteran suborders. Mol. Cells 28, 155–165.

Lent, H., Wygodzinsky, P.W., 1979. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. Bulletin of the AMNH; v. 163, article 3.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Gao, J., Cai, W., 2013. Complete mitochondrial genome of the assassin bug *Oncocephalus breviscutum* (Hemiptera: Reduviidae). Mitochondrial DNA 0,1–2.

Li, H., Gao, J., Liu, H., Liu, H., Liang, A., Zhou, X., Cai, W., 2011. The architecture and complete sequence of mitochondrial genome of an assassin bug *Agriosphodrus dohrni* (Hemiptera: Reduviidae). Int. J. Biol. Sci. 7, 792.

Li, H., Liu, H., Song, F., Shi, A., Zhou, X., Cai, W., 2012a. Comparative mitogenomic analysis of damsel bugs representing three tribes in the family Nabidae (Insecta: Hemiptera). PloS One 7, e45925.

Li, H., Liu, H., Shi, A., Štys, P., Zhou, X., Cai, W., 2012b. The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). PloS One 7, e29419.

Li, M., Tian, Y., Zhao, Y., Bu, W., 2012c. Higher Level Phylogeny and the First Divergence Time Estimation of Heteroptera (Insecta: Hemiptera) Based on Multiple Genes. PLoS ONE 7, e32152.

Li, H., Shi, A., Song, F., Cai, W., 2014. Complete mitochondrial genome of the flat bug *Brachyrhynchus hsiaoi* (Hemiptera: Aradidae). Mitochondrial DNA 0,1–2.

Lin, C.-P., Danforth, B.N., 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30,

Liu, L., Li, H., Song, F., Song, W., Dai, X., Chang, J., Cai, W., 2012. The mitochondrial genome of *Coridius chinensis* (Hemiptera: Dinidoridae). Zootaxa 3537, 29–40.

Macey, J.R., Schulte, J.A., Larson, A., Papenfuss, T.J., 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol. Biol. Evol. 15, 71–75.

Mahner, M., 1993. Systema Cryptoceratorum Phylogeneticum (Insecta, Heteroptera). Stuttgart, Germany: Eds. Schweizerbart.

Maldonado Capriles, J., 1990. Systematic catalogue of the Reduviidae of the world (Insecta: Heteroptera). Caribb. J. Sci. Spec. Ed.

Malé, P.-J.G., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., Lhuillier, E., Scotti-Saintagne, C., Tinaut, A., Chave, J., 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol. Ecol. Resour.

Mariette, J., Escudié, F., Allias, N., Salin, G., Noirot, C., Thomas, S., Klopp, C., 2012. NG6: Integrated next generation sequencing storage and processing environment. BMC Genomics 13, 462.

Moritz, C., Brown, W.M., 1986. Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233, 1425–1427.

Moritz, C., Dowling, T.E., Brown, W.M., 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 269–292.

Nagaike, T., 2005. Human Mitochondrial mRNAs Are Stabilized with Polyadenylation Regulated by Mitochondria-specific Poly(A) Polymerase and Polynucleotide Phosphorylase. J. Biol. Chem. 280, 19721–19727.

Nguyen, V.H., Lavenier, D., 2009. PLAST: parallel local alignment search tool for database comparison. BMC Bioinformatics 10, 329.

Ojala, D., Montoya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature 290, 470-474.

Oliveira, M.T., Barau, J.G., Junqueira, A.C.M., Feijão, P.C., Rosa, A.C. da, Abreu, C.F., Azeredo-Espin, A.M.L., Lessinger, A.C., 2008. Structure and evolution of the mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: the Muscidae (Diptera: Calyptratae) perspective. Mol. Phylogenet. Evol. 48, 850–857.

Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353–358.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Saito, S., 2005. Replication origin of mitochondrial DNA in insects. Genetics 171, 1695–1705.

Salvato, P., Simonato, M., Battisti, A., Negrisolo, E., 2008. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). BMC Genomics 9, 331.

Schuh, R.T., Weirauch, C., Wheeler, W.C., 2009. Phylogenetic relationships within the Cimicomorpha (Hemiptera: Heteroptera): a total-evidence analysis. Syst. Entomol. 34, 15–48.

Shao, R., Barker, S.C., 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. Mol. Biol. Evol. 20, 362–370.

Shao, R., Barker, S.C., Mitani, H., Takahashi, M., Fukunaga, M., 2006. Molecular mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus *Leptotrombidium* (Acari: Acariformes). J. Mol. Evol. 63, 251–261.

Shi, A.M., Li, H., Bai, X.S., Dai, X., Chang, J., Guilbert, E., Cai, W., 2012. The complete mitochondrial genome of the flat bug *Aradacanthia heissi* (Hemiptera: Aradidae). Zootaxa 3238, 23–38.

Signorovitch, A.Y., Buss, L.W., Dellaporta, S.L., 2007. Comparative Genomics of Large Mitochondria in Placozoans. PLoS Genet. 3, e13.

Som, A., 2014. Causes, consequences and solutions of phylogenetic incongruence. Brief. Bioinform. doi:10.1093/bib/bbu015

Song, N., Liang, A.-P., 2009. Complete Mitochondrial Genome of the Small Brown Planthopper, *Laodelphax striatellus* (Delphacidae: Hemiptera), with a Novel Gene Order. Zoolog. Sci. 26, 851–860.

Song, N., Liang, A.-P., Bu, C.-P., 2012. A molecular phylogeny of Hemiptera inferred from mitochondrial genome sequences. PloS One 7, e48778.

Song, W., Li, H., Song, F., Liu, L., Wang, P., Xun, H., Dai, X., Chang, J., Cai, W., 2013. The complete mitochondrial genome of a tessaratomid bug, *Eusthenes cupreus* (Hemiptera: Heteroptera: Pentatomomorpha: Tessaratomidae). Zootaxa 3620, 260–272.

Stamatakis, A., 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics btu033.

Stewart, J.B., Beckenbach, A.T., 2009. Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. Gene 445, 49–57.

Straub, S.C., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am. J. Bot. 99, 349–364.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Thao, M.L., Baumann, L., Baumann, P., 2004. Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). BMC Evol. Biol. 4, 25.

Thompson, K.F., Patel, S., Williams, L., Tsai, P., Constantine, R., Baker, C.S., Millar, C.D., 2014. High coverage of the complete mitochondrial genome of the rare Gray's beaked whale (*Mesoplodon grayi*) using Illumina next generation sequencing. Mitochondrial DNA 0,1–2.

Tian, Y., Zhu, W., Li, M., Xie, Q., Bu, W., 2008. Influence of data conflict and molecular phylogeny of major clades in Cimicomorphan true bugs (Insecta: Hemiptera: Heteroptera). Mol. Phylogenet. Evol. 47, 581–597.

Veale, A.J., Williams, L., Tsai, P., Thakur, V., Lavery, S., 2014. The complete mitochondrial genomes of two chiton species (*Sypharochiton pelliserpentis* and *Sypharochiton sinclairi*) obtained using Illumina next generation sequencing. Mitochondrial DNA 0,1–2.

Walberg, M.W., Clayton, D.A., 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. 9, 5411–5421.

Wang, P., Li, H., Wang, Y., Zhang, J.-H., Dai, X., Chang, J., Hu, B.-W., Cai, W.-Z., 2013. The mitochondrial genome of the plant bug *Apolygus lucorum* (Hemiptera: Miridae): Presently known as the smallest in Heteroptera. Insect Sci. 21, 159-173.

Wang, Y., Li, H., Xun, H., Cai, W., 2014. Complete mitochondrial genome sequence of the plant bug *Adelphocoris fasciaticollis* (Hemiptera: Heteroptera: Miridae). Mitochondrial DNA 0,1–2.

Weirauch, C., 2008. Cladistic analysis of Reduviidae (Heteroptera: Cimicomorpha) based on morphological characters. Syst. Entomol. 33, 229–274.

Weirauch, C., Munro, J.B., 2009. Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae), based on mitochondrial and nuclear ribosomal genes. Mol. Phylogenet. Evol. 53, 287–299.

Weirauch, C., Schuh, R.T., 2011. Systematics and Evolution of Heteroptera: 25 Years of Progress. Annu. Rev. Entomol. 56, 487–510.

Wheeler, A.G., 1997. Zoological Catalogue of Australia 27.3 A: Hemiptera: Heteroptera (Coleorrhyncha to Cimicomorpha). Ann. Entomol. Soc. Am. 90, 703–704.

Wheeler, W.C., Schuh, R.T., Bang, R., 1993. Cladistic relationships among higher groups of Heteroptera: congruence between morphological and molecular data sets. Insect Syst. Evol. 24, 121–137.

Wolstenholme, D.R., 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141, 173–216.

Xie, Q., Tian, Y., Zheng, L., Bu, W., 2008. 18S rRNA hyper-elongation and the phylogeny of Euhemiptera (Insecta: Hemiptera). Mol. Phylogenet. Evol. 47, 463–471.

Yang, W., Yu, W., Du, Y., 2013. The complete mitochondrial genome of the sycamore lace bug *Corythucha ciliata* (Hemiptera: Tingidae). Gene 532, 27–40.

Yu, S., Wang, Y., Rédei, D., Xie, Q., Bu, W., 2013. Secondary structure models of 18S and 28S rRNAs of the true bugs based on complete rDNA sequences of *Eurydema maracandica* Oshanin, 1871 (Heteroptera, Pentatomidae). ZooKeys 319, 363–377.

Yukuhiro, K., Sezutsu, H., Itoh, M., Shimizu, K., Banno, Y., 2002. Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild Mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. Mol. Biol. Evol. 19, 1385–1389.

Yuting, D., Li, Jiang, P., Song, F., Ye, Z., Yuang, X., Dai, X., Chang, J., Cai, W., 2012. Sequence and organization of the mitochondrial genome of an urostylidid bug, *Urochela quadrinotata* Reuter (Hemiptera: Urostylididae). Entomotaxonomia 34, 613–623.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.

Zhang, D.-X., Hewitt, G.M., 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem. Syst. Ecol. 25, 99–120.

Zhang, D.-X., Szymura, J.M., Hewitt, G.M., 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. J. Mol. Evol. 40, 382–391.

Zhang, Q.-L., Yuan, M.-L., Shen, Y.-Y., 2013. The complete mitochondrial genome of *Dolycoris baccarum* (Insecta: Hemiptera: Pentatomidae). Mitochondrial DNA 24, 469–471.

Zhou, Z., Huang, Y., Shi, F., 2007. The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A+T-rich region of 70 bp in length. Genome Natl. Res. Counc. Can. 50, 855–866.

APPENDICES

Apprendix 1: DNA extraction protocol

1. Place the whole specimen (or tissue) in a 1.5 mL tube with 800 µL TE9 buffer (0.5 M Tris, pH 9.0, 20 mM EDTA, 10 mM NaCl). Incubate at 56 °C for 1 hour in a shaky incubator (more for bigger or thicker specimens).
2. Pipette and discard the 800 uL of TE9 buffer and add 180 uL of PBS buffer and homogenize the sample using a disposable pestle. Clean the pestle with 95% EtOH and water if reused.
3. Add 20 uL of proteinase K and 200 uL of buffer AL. Mix thoroughly by vortexing and incubate at 56°C for 1-3 hours. Make sure all the sample is submerged into the solution.
4. Add 240 uL of 100% ethanol (chilled at 4°C in the fridge) to the sample, and mix thoroughly by vortexing.
5. Pipette the mixture from the previous step (including any precipitate) into the DNeasy spin column placed in a 2 mL collection tube. Centrifuge at >6000g (8000 rpm) for 1 minute. Discard flow-through and collection tube.
6. Place the spin column in a new 2 mL collection tube, add 500 uL of buffer AW1 and centrifuge for 1 minute at >6000g (8000 rpm). Discard flow-through and collection tube.
7. Place de column in a new 2 mL collection tube, add 500 uL of buffer AW2 and centrifuge for 3 minutes at 11000 g (14000 rpm) to dry the membrane. Discard flow-through and reuse the collection tube.
8. Centrifuge for 1 min at at 11000 g (14000 rpm) to dry the membrane. Discard flow-through and the collection tube.
9. Place the column in 1.5 mL tube. Leave 3-5 min at room temperature to make sure all the ethanol is evaporated.
10. Add 120 uL of elution buffer AE onto the membrane. Leave for 5 minute at room temperature to make sure the buffer is absorbed by the membrane.
11. Centrifuge for 1 minute at 6000 g (8000 rpm) to elute.

Appendix 2: Coverage analysis for the inference of duplication copy number.

We used the coverage information to infer the copy number of duplicated elements, assuming that the read coverage of a particular genomic region is proportional to its copy fraction in the sample. We compared the number of reads mapping on a DNA fragment present only once in the mitogenome to that of DNA fragment belonging to a duplicated element present in two copies at least. Coverage values at two nucleotide positions are more likely to be similar when the distance between them is decreasing. Therefore, we based our comparisons on short and close fragments in order to minimize bias due to variations in coverage along the mitogenome sequence.

We compared the number of Illumina reads perfectly mapping on the first 30 bp/the last 30 bp of the duplicated element (present at least twice in the mitogenome) to the number of reads perfectly mapping on the 30 bp preceding the first copy/following the last copy of the duplicated element respectively (present only once in the mitogenome). The average of the two ratios was rounded to the nearest integer in order to infer the number of copies.

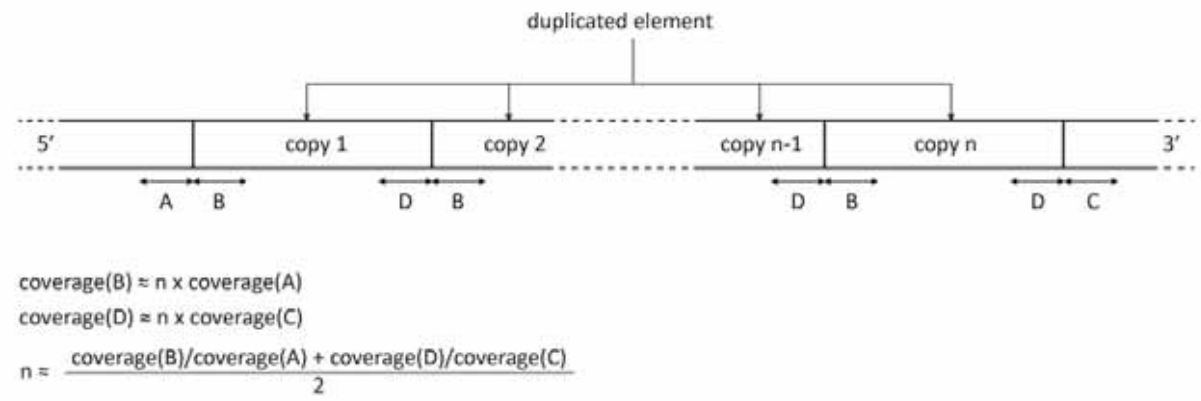


Figure: Schematization of the coverage analysis approach to estimate the copy number of repeated elements.

Table: Results of the coverage analysis: comparison of the number of reads perfectly mapping on the 30 bp preceding the first copy (A)/following the last copy (C) of the repeated element to that of the the first 30 bp (B)/the last 30 bp (D) of the repeated element.

Location of the first repeat	Length of the repeated element	A	B	C	D	Average ratio: (A/B+C/D)/2	Inferred number of copy*
5,845	131	148	279	172	259	1,69	2
14,452	857	167	310	58	153	2,24	2
16,200	74	55	317	90	354	4,84	5

*Nearest integer of the average ratio

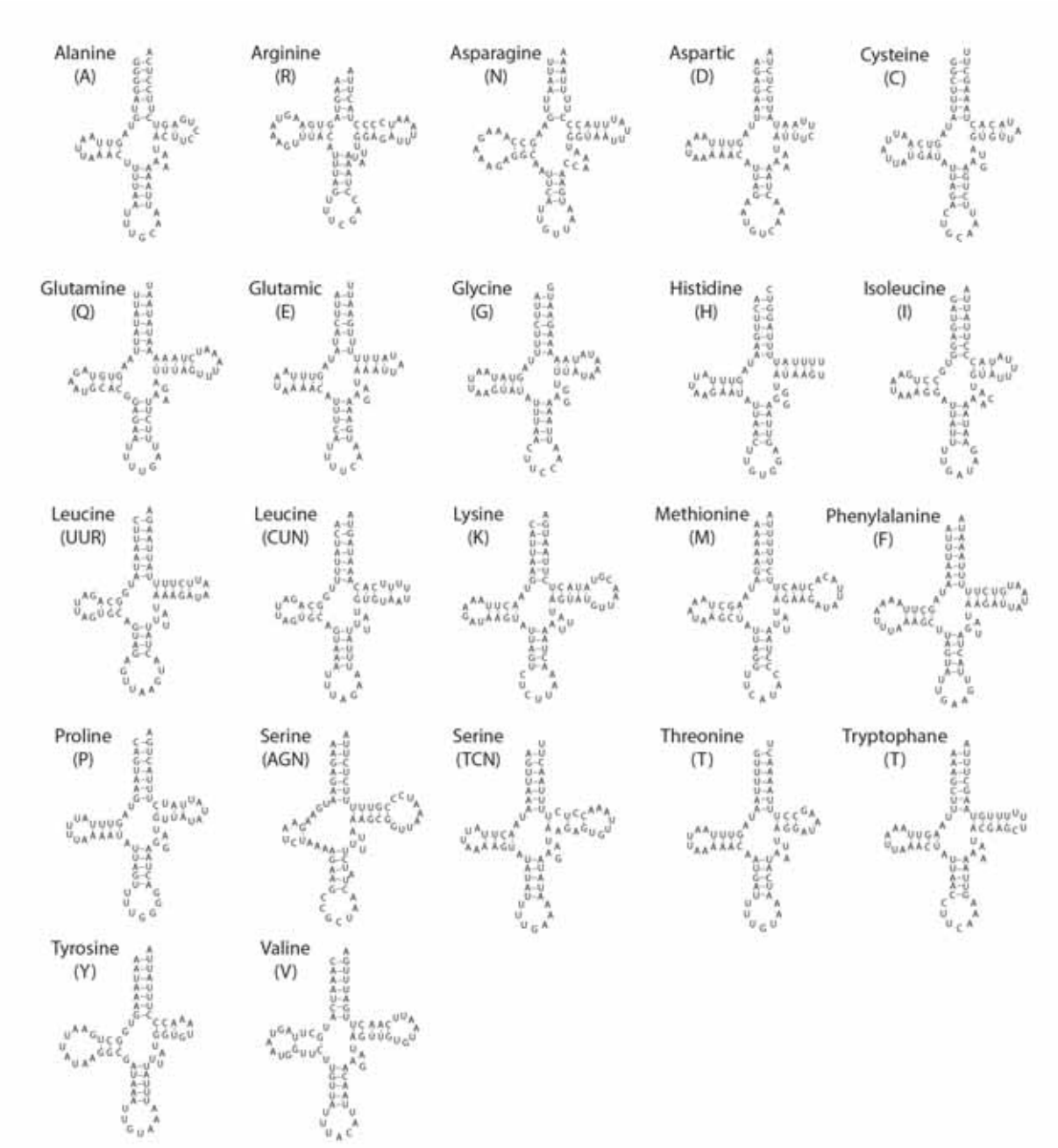
Appendix 3: Nucleotide composition of the mitogenomes of the following assassin bugs: Brontostoma colossus (B), Agriosphodrus dohrni (A), Oncocephalus breviscutum (B), Sirthenea flavipes (S), Triatoma dimidiate (T) and Valentia hoffmanni (V).

	A+T%						AT skew						GC skew					
	B	*A*	*O*	*S*	*T*	*V*	*B*	*A*	*O*	*S*	*T*	*V*	*B*	*A*	*O*	*S*	*T*	*V*
Whole	73.5	72.2	80.3	78.8	69.5	73.8	0.16	0.08	0.22	0.02	0.16	0.12	-0.22	-0.12	-0.16	-0.41	-0.27	-0.21
PCGs	72.8	71.7	74.4	71.3	68.8	73.4	-0.14	-0.14	-0.15	-0.15	-0.17	-0.16	-0.01	-0.01	0.03	0.00	-0.02	0.00
1st CP	69.4	73.5	68.4	66.6	69.2	74.6	-0.01	-0.09	-0.02	0.00	-0.13	-0.10	0.18	0.01	0.21	0.19	-0.03	0.01
2nd CP	66.6	70	66.1	66	68	72.3	-0.39	-0.18	-0.39	-0.41	-0.24	-0.20	-0.09	0.01	-0.10	-0.11	0.03	0.02
3rd CP	82.8	71.5	89	81.3	69.3	73.4	-0.04	-0.16	-0.08	-0.07	-0.17	-0.16	-0.20	-0.04	-0.09	-0.13	-0.08	-0.02
tRNAs	76.8	73.4	75.3	74.1	74.3	76	-0.01	0.02	0.04	0.03	0.03	0.01	0.17	0.11	0.14	0.16	0.15	0.14
rRNAs	77.1	74.8	75.1	73.3	73.1	75.7	-0.13	-0.09	-0.14	-0.16	-0.18	-0.11	0.31	0.27	0.28	0.27	0.31	0.30
Control region	70.8	71.9	71.8	70.3	66	69.9	0.08	0.05	0.06	0.13	0.20	0.03	-0.15	-0.17	-0.15	-0.16	-0.29	-0.25

Appendix 4: PCGs start and stop codons in the mitogenome of six assassin bugs.

Gene	**Start codon**						**Stop codon**					
	B	*A*	*O*	*S*	*T*	*V*	*B*	*A*	*O*	*S*	*T*	*V*
ND2	ATT	ATT	ATT	ATT	ATC	ATT	T--	TAA	TAA	TAA	TAG	TAA
COXI	ATG	ATG	ATG	ATG	ATG	ATG	T--	T--	T--	TAA	T--	T--
COXII	ATT	ATC	ATG	ATA	ATA	ATC	T--	TAA	T--	TAA	T--	T--
ATP8	ATT	ATT	ATA	ATA	ATA	ATT	TAA	TAA	TAA	TAA	T--	TAA
ATP6	ATG	ATG	ATG	ATG	ATG	ATG	TAA	TAA	TAG	TAA	TAA	TAG
COXIII	ATG	ATG	ATG	ATG	ATG	ATG	TAA	T--	T--	TAG	TAA	T--
ND3	ATA	ATA	ATT	ATA	ATA	ATA	TAA	TAA	TA-	T--	T--	T--
ND5	ATT	ATG	ATT	ATT	GTG	ATG	TA-	T--	T--	T--	T--	T--
ND4	ATG	ATG	ATG	ATG	ATG	ATG	TAA	TAA	TAG	T--	T--	TAA
ND4L	ATT	ATG	GTG	GTG	ATG	ATG	TAA	TAA	TAA	TAA	TAA	TAA
ND6	ATA	ATG	ATG	ATT	ATA	ATG	TAA	TAA	TAA	TAA	TAA	TAA
CYTB	ATG	ATG	ATA	ATG	ATG	ATG	TA-	TAA	T--	TAA	TAA	TAG
ND1	ATT	GTG	ATG	GTG	ATA	GTG	TAA	TAA	TAG	TAG	T--	TAA

Appendix 5: Secondary structure of Brontostoma colossus mitochondrial tRNAs.



Appendix 6: Codon usage of Brontostoma colossus mitogenome protein-coding genes.

Amino Acid	Codon	N	RSCU	Amino Acid	Codon	N	RSCU
Ala (A)	GCA	55	1.86	Phe (P)	CCA	51	1.47
	GCC	12	0.41		CCC	20	0.58
	GCG	3	0.10		CCG	5	0.14
	GCT	48	1.63		CCT	63	1.81
Cys (C)	TGC	6	0.19	Gln (Q)	CAA	47	1.68
	TGT	56	1.81		CAG	9	0.32
Asp (D)	GAC	15	0.43	Arg (R)	CGA	25	2.04
	GAT	54	1.57		CGC	4	0.33
Glu (E)	GAA	58	1.66		CGG	6	0.49
	GAG	12	0.34		CGT	14	1.14
Phe (F)	TTC	82	0.46	Ser (S)	AGA	72	1.66
	TTT	276	1.54		AGC	9	0.21
Gly (G)	GGA	111	1.93		AGG	8	0.18
	GGC	19	0.33		AGT	34	0.78
	GGG	31	0.54		TCA	85	1.96
	GGT	69	1.20		TCC	22	0.51
His (H)	CAC	22	0.57		TCG	9	0.21
	CAT	55	1.43		TCT	108	2.49
Ile (I)	ATC	56	0.29	Tyr (T)	ACA	101	2.10
	ATT	328	1.71		ACC	23	0.48
Lys (K)	AAA	92	1.72		ACG	7	0.15
	AAG	15	0.28		ACT	61	1.27
Leu (L)	CTA	68	0.80	Val (V)	GTA	57	1.37
	CTC	10	0.12		GTC	13	0.31
	CTG	10	0.12		GTG	7	0.17
	CTT	54	0.63		GTT	90	2.16
	TTA	295	3.46	Trp (W)	TGA	78	1.66
	TTG	74	0.87		TGG	16	0.34
Met (M)	ATA	247	1.70	Tyr (Y)	TAC	29	0.33
	ATG	43	0.30		TAT	145	1.67
Asn (N)	AAC	38	0.46				
	AAT	128	1.54				

N: total number in all PCGs, RSCU: relative synonymous codon usage

ARTICLE 2:

COMPLETE MITOCHONDRIAL GENOME OF *LUTZOMYIA* (*NYSSOMYIA*) *UMBRATILIS*, THE MAIN VECTOR OF *LEISHMANIA GUYANENSIS*

Arthur Kocher[1,2], Jean-Charles Gantier[3], Hélène Holota[1], Céline Jeziorski[4,5], Eric Coissac[6], Anne-Laure Bañuls[2], Romain Girod[7], Pascal Gaborit[7] and Jérôme Murienne[1*]

Published in *Mitochondrial DNA* (doi: 10.3109/19401736.2015.1022748)

* [1]CNRS, Université Toulouse III Paul Sabatier, ENFA; UMR5174 EDB (Laboratoire Evolution et Diversité Biologique), Toulouse, France, [2]UMR MIVEGEC (IRD 224 – CNRS 5290), Université de Montpellier), Montpellier, France, [3]Laboratoire des Identifications Fongiques et Entomo-parasitologiques, Mennecy, France, [4]INRA, UAR 1209 Département de génétique animale, INRA Auzeville, Castanet-Tolosan, France, [5]GeT–PlaGe, Genotoul, INRA Auzeville, Castanet-Tolosan, France, [6]Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université´ Joseph Fourier, Grenoble, France, and [7]Medical Entomology Unit, Institut Pasteur de la Guyane, Cayenne Cedex, French Guiana

ABSTRACT

The nearly complete mitochondrial genome of *Lutzomyia umbratilis* Ward & Fraiha, 1977 (Psychodidae: Phlebotominae), considered as the main vector of *Leishmania guyanensis*, is presented. The sequencing has been performed on an Illumina Hiseq 2500 platform, with a genome skimming strategy. The full nuclear ribosomal RNA segment was also assembled. The mitogenome of *L. umbratilis* was determined to be at least 15,717 bp-long and presents an architecture found in many mitogenomes of insect (13 protein-coding genes, 22 transfer RNAs, two ribosomal RNAs, and one non-coding region also referred as the control region). The control region contains a large repeated element of c. 370 bp and a poly-AT region of unknown length. This is the first mitogenome of Psychodidae to be described.

MITOGENOME ANNOUNCEMENT

Lutzomyia sensu lato is a New World genus of sand flies (Psychodidae: Phlebotominae) that currently includes 477 described species. It comprises all the proven vector species of Leishmaniasis in the New World (Killick-Kendrick, 1990). *Lutzomyia umbratilis* (Ward & Fraiha 1977) is distributed in several countries of northern South America (Scarpassa & Alencar, 2012) and is considered to be the main vector of *Leishmania guyanensis* (Pajot et al., 1986). Here, the complete mitochondrial genome of *L. umbratilis* was sequenced using a genome-skimming approach (Cally *et al.*, 2014; Kocher *et al.*, 2014, 2015). This is the first mitogenome of Psychodidae to be described.

Specimens of *L. umbratilis* were collected using CDC light traps in the surrounding of Saint-Georges de l'Oyapock, French Guiana on 2 October 2007, processed in the Pasteur institute in Cayenne and stored at the LIFE laboratory, Mennecy, France. After traditional morphological identification, heads, wings, and genitalia were kept as vouchers (IDs: 1664B, 1667B, and 1669B). Total genomic DNA was extracted from the thoraxes using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA). DNA was then sent for library construction and sequencing to the GeTPlaGe core facilities of Genotoul (Toulouse, France). The library was constructed with 162 ng of DNA using the Illumina TruSeq Nano DNA Sample Prep Kit (Illumina Inc. San Diego, CA). The library was hybridized on 1/24th of a lane of Illumina Hiseq 2500 flow cell using the Illumina TruSeq PE Cluster Kit v.3 (Illumina Inc., San Diego, CA). Paired-end reads of 100 nucleotides were collected using the Illumina TruSeq SBS Kit v.3 (Illumina Inc., San Diego, CA) (200 cycles). The mitochondrial genome and the nuclear ribosomal gene segment were assembled using a genome walking approach and annotated as previously described (Kocher et al., 2014). The sequences were deposited in GenBank under the accession numbers KP702938 and KP702939. Because this is the first mitogenome of Psychodidae available, we validated our sequence based on the 5′ region of the mitochondrial gene COI using the Barcoding of Life Database Identification Engine (http://www.boldsystems.org, Figure 1).

These results demonstrate that for a relatively small cost (below 280 euros), it is possible to efficiently assemble a complete sand fly mitogenome with sufficient coverage (the mean coverage in the coding region was 84x with a minimum of 30x) as well as a complete nuclear ribosomal cluster (the mean coverage was 101x with a minimum of 31x). The complete mitogenome of L. umbratilis is determined to be at least 15,717 bp long and presents an organization shared with many insects (Boore, 1999). The typical set of 37 genes

(13 protein-coding genes, 22 transfer RNA genes, and two ribosomal RNA genes) and one control region was identified. All protein-coding genes started with an ATN codon. Nine TAA and one TAG stop codons were identified. Three incomplete stop codons (TA or T) were found adjacent to transfer RNAs encoded on the same strand, consistent with the punctuation model for primary transcripts processing followed by 3' polyadenylation of mature mRNA (Nagaike, 2005; Ojala et al., 1981). The sequence of the control region was not completely assembled, but was determined to be at least 882 bp long. It contained at least two almost identical tandem-repeats of a ca 370 bp-long element and a poly-AT region of unknown length. The latter is probably related with a very low sequencing coverage of the region and the inability to complete the assembly.

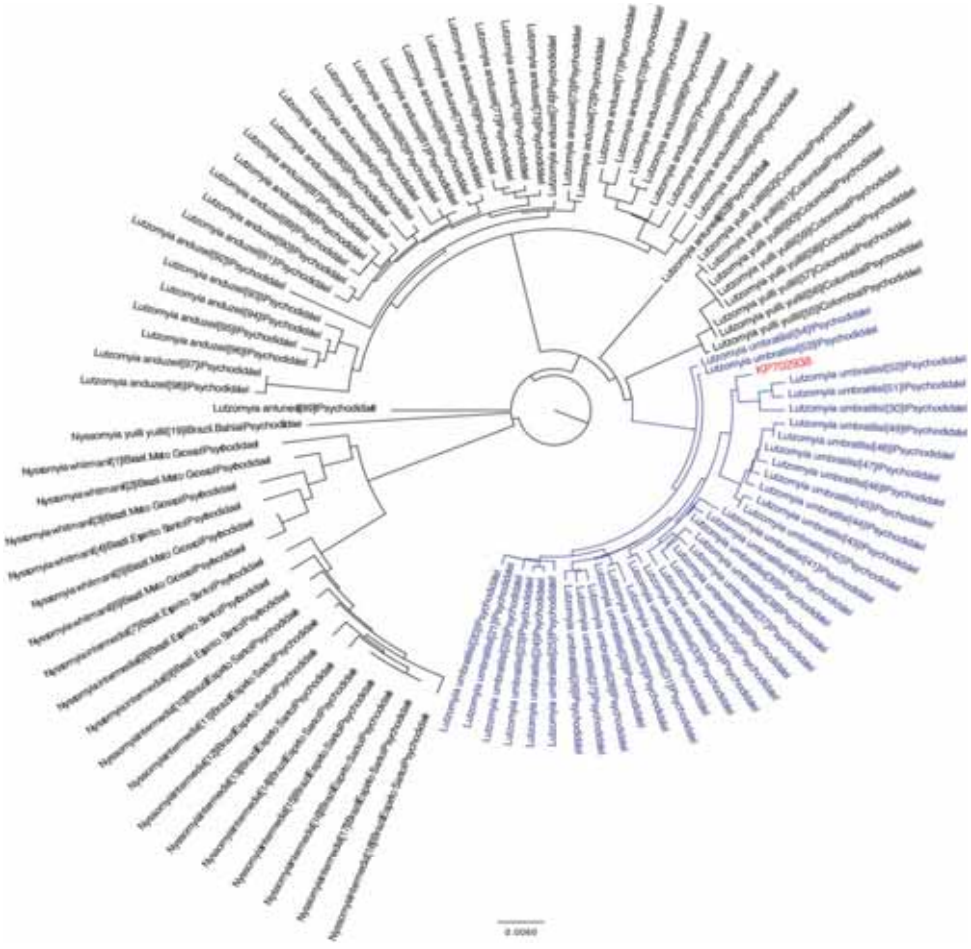


Figure 1: Neighbor-joining tree provided by the BOLD Identification Engine for the validation of our sample. The tree is based on the 5' region of the mitochondrial gene COI, which is the typical DNA barcode for animals. Our sequence is annotated with its GenBank accession number.

ACKNOWLEDGEMENTS

We would like to thank Romuald Carinci and Jean Issaly for their involvement in the collection and processing of the samples.

DECLARATION OF INTEREST

This work was supported by ‘‘Investissement d’Avenir’’ grants managed by Agence Nationale de la Recherche (CEBA, ANR-10-LABX-25-01; TULIP, ANR-10-LABX-41, and ANR-11-IDEX-0002-02) as well as project METABAR (ANR-11-BSV7-0020). The collection of sand flies was performed in the framework of the French National Project ANR 06- SEST-20 IAEL. The authors report that they have no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

Boore JL. (1999). Animal mitochondrial genomes. Nucleic Acids Res 27:1767–80.

Cally S, Lhuillier E, Iribar A, Garzon-Orduña I, Coissac E, Murienne J. (2014). Shotgun assembly of the complete mitochondrial genome of the neotropical cracker butterfly *Hamadryas epinome*. Mitochondrial DNA. (in press).

Killick-Kendrick R. (1990). Phlebotomine vectors of the leishmaniases: A review. Med Vet Entomol 4:1–24.

Kocher A, Guilbert E, Lhuillier E, Murienne J. (2015). Sequencing of the mitochondrial genome of the avocado lace bug *Pseudacysta perseae* (Heteroptera, Tingidae) using a genome skimming approach. C R Biol 338:49–160.

Kocher A, Kamilari M, Lhuillier E, Coissac E, Péneau J, Chave J, Murienne J. (2014). Shotgun assembly of the assassin bug *Brontostoma colossus* mitochondrial genome (Heteroptera, Reduviidae). Gene 552: 184–94.

Nagaike T. (2005). Human mitochondrial mRNAs are stabilized with polyadenylation regulated by mitochondria-specific poly(A) polymerase and polynucleotide phosphorylase. J Biol Chem 280: 19721–7.

Ojala D, Montoya J, Attardi G. (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–4.

Pajot FX, Chippaux JP, Geoffroy B, Dedet JP. (1986). La leishmaniose en Guyane française. Fluctuations saisonnières de la densité et du taux d’infection naturelle de *Lutzomyia umbratilis* Ward et Fraiha, 1977 en forêt dégradée. Cah ORSTOM Série Entomol Méd Parasitol 24:191–8.

Scarpassa VM, Alencar RB. (2012). *Lutzomyia umbratilis*, the main vector of *Leishmania guyanensis*, represents a novel species complex? PLoS One 7:e37341.

Ward RD, Fraiha H. (1977). *Lutzomyia umbratilis*, a new species of sandfly from Brazil (Diptera: Psychodidae). J Med Entomol 13: 313–17.

ARTICLE 3:
PHYLOGENOMICS OF NEW WORLD SAND FLIES (DIPTERA, PSYCHODIDAE) USING A *GENOME SKIMMING* APPROACH

Arthur Kocher[1,2], Pieternel Verschuren[1], Jean-Charles Gantier[3], Céline Jeziorski[4], Pascal Gaborit[5], Romain Girod[5], Anne-Laure Bañuls[2] and Jérôme Murienne[1].*

Submitted

---

* [1]CNRS, Université Toulouse III Paul Sabatier, ENFA; UMR5174 EDB (Laboratoire Evolution et Diversité Biologique), Toulouse, France; [2]UMR MIVEGEC (IRD 224 – CNRS 5290 – Université de Montpellier), 911 Avenue Agropolis, F34394 Montpellier, France; [3]Laboratoire des Identifications Fongiques et Entomo-parasitologiques, Mennecy, France; [4]GeT–PlaGe, Genotoul, INRA Auzeville, 31326 Castanet-Tolosan, France; [5]Medical Entomology Unit, Institut Pasteur de la Guyane, 23 Avenue Pasteur, BP 6010, 97306 Cayenne Cedex, French Guiana.

ABSTRACT

Phlebotominae (sand flies) are hematophagous dipterans responsible for the transmission of leishmaniasis worldwide. Despite the medical importance of the group, its evolutionary history is poorly understood. In the New World, conflicting sand fly classifications coexist, creating confusion in the literature, and few molecular studies have assessed the current taxonomy. Here, we used a genome skimming approach for phylogenetic reconstruction of Phlebotominae. We assembled the complete mitochondrial genomes and nuclear ribosomal RNA genes of 13 neotropical sand fly species using shotgun sequencing on an Illumina Hiseq platform. This dataset corresponds to *c*. 20,000 aligned nucleotides with a 100% completion across genes and ingroup taxa, which we further completed with three previously sequenced sand fly mitogenomes. The results of the phylogenetic analysis are robust with only few topological incongruences between nuclear and mitochondrial markers. The topology generally supports Galati's classification, but refutes the paraphyly of the genus *Lutzomyia sensu* Young & Duncan. We argue that the application of this strategy to a denser taxonomic sampling could provide solid knowledge of the evolutionary history of the group. We also discuss the side benefits of genome skimming for generating reference data for high-throughput bulk species identification methods.

INTRODUCTION

Sand flies (Diptera, Psychodidae, Phlebotominae) are a subfamily of hematophagous insects comprising close to 1000 species distributed worldwide (Galati, 2014). They are responsible for the transmission of leishmaniases, a group of parasitic diseases that causes over one million cases and approximately 20,000 human deaths per year worldwide (Alvar *et al.*, 2012). They are also involved in the transmission of other viral and bacterial diseases (Depaquit *et al.*, 2010; Ready, 2013). The New World has more than half of the species described so far, including 56 proven or suspected vectors of leishmaniases (Maroli *et al.*, 2013). In this region, two conflicting classifications currently coexist (see Akhoundi *et al.*, 2016, for a recent review of the taxonomy).

Young & Duncan (1994) proposed a phenetic classification, mainly based on that of Lewis *et al.* (1977), in which New World sand flies are divided in three genera: *Lutzomyia*, *Brumptomyia*, and *Waryiela*. The genus *Lutzomyia* contains the vast majority of species and is subdivided into 15 subgenera and 11 species groups. A new classification based on the manual (*i.e.* non computer-aided) cladistic analysis of 88 morphological characters was later proposed (Galati, 1995), with few minor recent updates (Galati, 2014). Phlebotominae was divided into two tribes, Hertigini and Phlebotomina, the latter divided again into 6 subtribes: Phlebotomina (*Phlebotomus*), Australophlebotomina, Brumptomyiina, Sergentomyiina, Lutzomyiina and Psychodopygina (Fig. 1). The subgenera of *Lutzomyia sensu* Young & Duncan were elevated to the generic rank and distributed among the subtribes Lutzomyiina, Psychodopygina, but also Sergentomyiina in which Old World species of the genus *Sergentomyia* (*sensu* Young & Duncan) are also found. This implies the paraphyly of the still mainly used *Lutzomyia sensu* Young & Duncan and, more generally, of New World sand flies, as one of the most remarkable consequences of this revised classification. Currently, Young & Duncan's taxonomy is still largely used, whereas Galati's is favoured among South American taxonomists. Apart from generating confusion in the literature, as well as hampering data storage and retrieval, this lack of taxonomic consensus highlights a broader issue: despite the medical importance of these insects, the evolutionary history of New World sand flies is poorly known.

So far, only few molecular studies have been conducted to assess the phylogenetic relationships among New World sand flies. The first were based on single genes and included a very limited number of species (Lins *et al.*, 2002; Mazzoni *et al.*, 2002). Later, Beati *et al.* (2004) included a more representative taxonomic sample and based their analysis on both

mitochondrial and nuclear markers (12S and 28S rRNAs), but the datasets were not combined and the tree topology was not well resolved. Cohnstaedt *et al.* (2011) based their study on the COI gene and focused on the *Verrucarum* species group (*sensu* Young & Duncan). More recently, Grace-Lema *et al.*, (2015) provided a comprehensive analysis of the Phlebotominae relationships by gathering public molecular data. However, few markers were available for the New World species (only a fragment of the COI gene for most of them), and their results showed poor statistical support for this group. As a consequence, Galati's classification remains the most substantial phylogenetic hypothesis for New World sand flies and still lacks molecular validation.

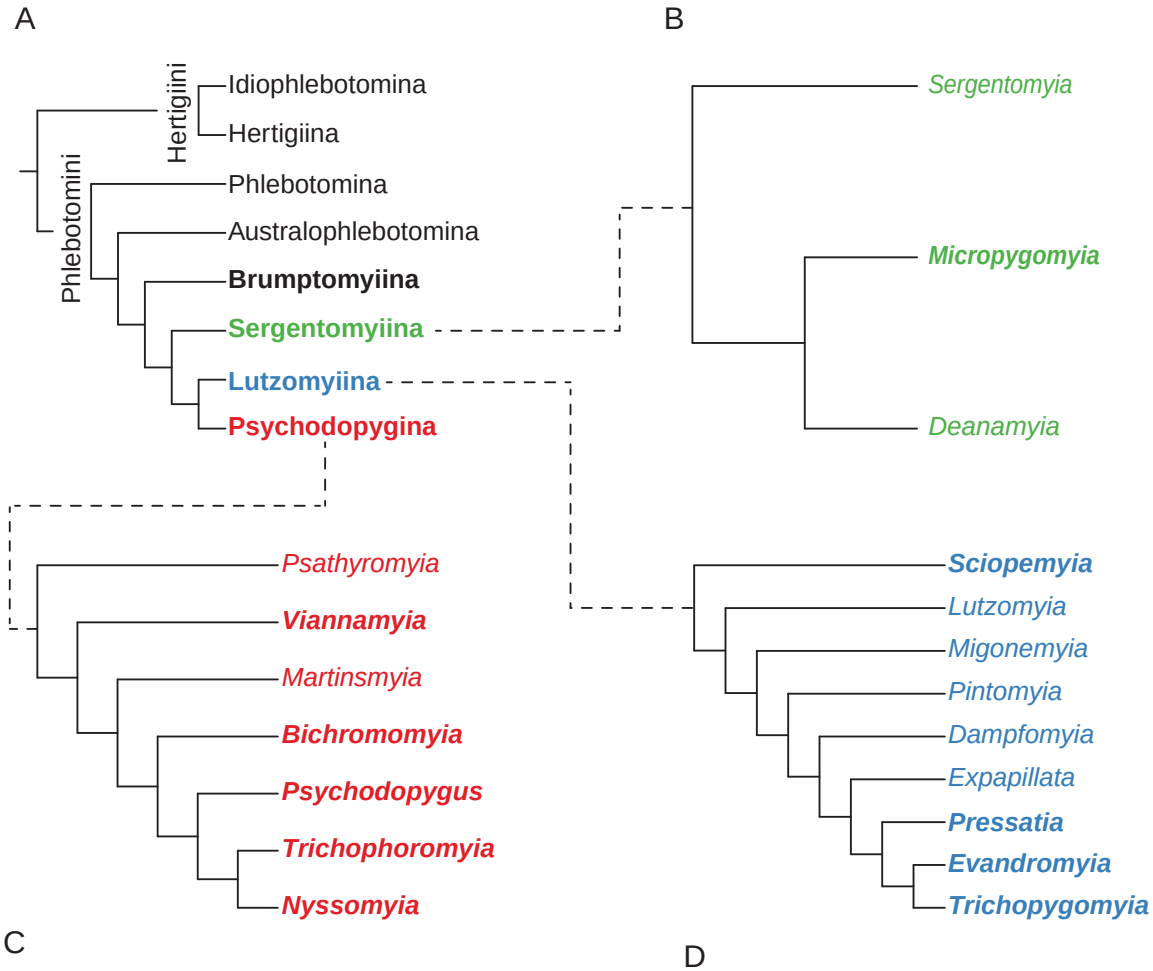


Figure 1: Simplified version of the classification proposed by Galati (1995). The clades represented in our dataset are in bold. **A:** Phylogenetic relationships between the different tribes and subtribes within Phlebotominae. **B, C, D:** Phylogenetic relationships between the different genera within the Sergentomyiina, Psychodopygina and Lutzomyiina subtribes.

In this study, we used a *genome skimming* approach to study the phylogeny of Neotropical sand flies. Originally developed for plant organelles, the genome skimming approach refers to the use of Next Generation Sequencing for retrieving the high-copy fractions of the genome with only little sequencing depth (Straub *et al.*, 2012). This strategy

has been successfully applied to assemble the complete mitochondrial genome and complete nuclear ribosomal cluster in many groups of insects (*e.g.* Kocher *et al.*, 2014, 2015) including sand flies (Kocher *et al.*, 2016a) and mitogenomic data has allowed robust phylogenetic reconstruction at the family level (e.g. Kim *et al.*, 2011; Cameron *et al.*, 2012). We assembled complete mitochondrial genomes and nuclear ribosomal RNA genes for 13 sand fly species representing 11 genera and 4 sub-tribes (*sensu* Galati). This molecular dataset was used together with three previously sequenced sand fly mitogenomes (Ye *et al.*, 2015; Kocher *et al.*, 2016a) for a phylogenetic analysis of New World sandflies. More specifically, we aim to test the monophyly of the neotropical species (i.e. test the monophyly of the genus *Lutzomyia sensu* Young & Duncan and to assess the validity of the latest classification scheme based on molecular data.

MATERIAL AND METHODS

Galati's classification, which is based on a cladistic analysis of morphological characters, currently constitutes the reference phylogenetic hypothesis for New World sand flies. It will be used throughout the text, but the alternative classification of Young & Duncan can be found in the Supplementary Material.

Laboratory procedures

Specimens of sand flies were collected with Center for Disease Control (CDC) light traps modified with LEDs in various locations in French Guiana and the Caribbean between 2007 and 2012 (see Supporting information for details). The samples were then processed at Institut Pasteur de la Guyane (Cayenne, French Guiana) and stored at the LIFE laboratory (Mennecy, France). After dissection, morphological identifications were performed based on heads, wings and genitalia, while thoraxes were kept for molecular study. Thirty-seven specimens were selected to represent 13 species from 11 distinct genera were selected. Total genomic DNA was extracted from the thoraxes using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA). Several specimens (up to three when possible) were pooled to obtain sufficient material for library construction. The quality and quantity of extracted genomic DNA were evaluated using a nanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and a PicoGreen double-stranded DNA quantitation assay kit (Life Technologies, Carlsbad, CA, USA).

The genomic DNA was then sent for library construction and sequencing to the GeTPlaGe core facilities of Genotoul (Toulouse, France). One library was constructed for

each species using the TruSeq Nano DNA Sample Prep Kit (Illumina Inc. San Diego, CA). The libraries were multiplexed with samples from other projects. The pool of libraries (24 libraries in total) was then hybridized on one lane of Illumina Hiseq 2500 flow cell using the TruSeq PE Cluster Kit v.3 (Illumina Inc., San Diego, CA). Paired-end reads of 150 nucleotides were collected using the TruSeq SBS Kit v.3 (Illumina Inc., San Diego, CA) (200 cycles). Quality filtering was performed with the Consensus Assessment of Sequence and Variation (CASAVA) pipeline. Sequence data were stored on the NG6 platform (Mariette *et al.*, 2012).

Sequence assembly and annotation

The Organelle.Assembler (available at pythonhosted.org/ORG.asm) was used to first assemble the reads matching the mitochondrial genome and nuclear rRNA genes of *Ny. umbratilis* (Kocher *et al.*, 2016a) and then performing a walking strategy by iteratively mapping the reads to the seeds until the complete sequences were obtained. If necessary, the software can efficiently take into account the paired-end information to fill the gaps using an additional alignment/assembly procedure.

The mitogenomes and nuclear ribosomal segment were then annotated by similarity with the genome of *Ny. umbratilis* using Geneious 9 Pro (Kearse *et al.*, 2012). The annotations of the protein-coding genes (PCGs) were adjusted by checking the open reading frames. The reads were remapped to the assembled mitogenomes and nuclear ribosomal genes using bowtie (Langmead *et al.*, 2009) and coverage statistics were computed using samtools (Li *et al.*, 2009).

Phylogenetic analyses

All previously sequenced Psychodidae mitogenomes were added to the newly generated dataset for phylogenetic analysis (*i.e. Ny. umbratilis, Phlebotomus chinensis* and *Ph. papatasi*; Kocher et al., 2015a; Ye et al., 2015). The 13 mitochondrial PCGs, both mitochondrial ribosomal RNA genes (12S and 16S rRNAs) and both nuclear ribosomal RNA genes (18S and 28S rRNAs) were used. The two Phlebotomus were used as outgroups. For the analysis of nuclear DNA only, *Brumptomyia travassosi* was used as outgroup because nuclear data were not available for *Phlebotomus*. The PCGs were aligned based on their corresponding amino acid sequences to ensure conservation of the reading frame using TranslatorX (Abascal, Zardoya, & Telford, 2010). Before back-translation, Gblocks (Castresana, 2000) was used to remove highly divergent regions. All ribosomal genes were

aligned separately using MAFFT 7 (Katoh & Standley, 2013) with the default parameter values, and divergent regions were cleaned using Gblocks with the "-a" option to remove all positions containing gaps. Sequences were then concatenated using FASconCAT-G (Kück & Meusemann, 2010), resulting in a total of 18,403 aligned nucleotides. The best partitioning schemes and substitution models were determined with PartitionFinder v.1.1.1 (Lanfear *et al.*, 2012) using the greedy algorithm to search for the best partitioning scheme and the Bayesian Information Criterion (BIC) for best scheme selection. One data block was defined for each gene at each codon position (for PCGs).

Saturation (multiple substitutions at the same position) can be a significant problem in phylogenetics, especially for mitochondrial DNA because of its rapid rate of evolution (Lin & Danforth, 2004; Jeffroy *et al.*, 2006). Following Jeffroy *et al.* (2006), we estimated the extant of saturation for each gene at each nucleotide position by estimating the slope of the scatterplot of uncorrected versus K80 corrected pairwise distances; the closest the slope is to one, the less saturated is the nucleotide dataset. The results indicate a low level of saturation compared to that mentioned in other studies (see for instance Jeffroy *et al.*, 2006; Comte, Murienne, & Grenouillet, 2014), even at the third codon positions of PCGs (slope between 0.46 and 0.60) which we therefore decided to keep for phylogenetic analyses.

To evaluate the potential incongruence between mitochondrial and nuclear loci, the following analyses were first performed on both datasets separately and then on the combined dataset. Maximum Likelihood (ML) analyses were performed with RAxML 8 (Stamatakis, 2014) using the best nucleotide substitution model and best partition scheme as identified by PartitonFinder. We used a rapid bootstrap procedure (Stamatakis, Hoover, & Rougemont, 2008) with an automated bootstopping option to evaluate the sufficient number of replicates. Analyses were performed on one lane of the EDB lab High Performance Computing system, comprising a dual-processor Intel Xeon E5-2770v3 (2.3 GHz, 12core) and 96 Gb of RAM. We used the hybrid version of RAxML with multithreading (option -T 8) and AVX vectorization for likelihood computation while distributing the bootstrap replicates over 6 instances of multiprocessing (mpiexec -by node -np 6, with the RAxML option --set-thread-affinity); thus using a total of 48 threads for the computation. Bayesian Inference (BI) was performed with MrBayes v.3.2.2.(Ronquist *et al.*, 2012). The same partitions and models were used as in the Maximum Likelihood analysis and all partitions were allowed to evolve under different rates. Coupled MCMC were run for 2,000,000 generations in four independent runs, each with 10 chains running simultaneously. Trees were sampled every 5,000 iterations. The 50% majority rule tree was selected after a relative burning of 25%.

RESULTS

Genome organization and composition

The complete mitochondrial genomes (excluding the control regions) and nuclear rRNA genes (18S, 28S and 5.8S) were successfully assembled for the 13 species. Only a *c.* 60 bp portion at the 5' end of the 12S rRNA gene of *Micropygomyia chassigneti* was not retrieved. We were not able to unambiguously assemble the Internal transcribed spacers (ITS1 and 2) for all species, probably because of some variability between specimens of the same species pooled in the samples. All mitogenomes have the same architecture and orientation as those of the two species of *Phlebotomus*, which is shared with many insects and crustaceans (Boore et al., 1998). They are composed of 37 genes containing 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, two ribosomal RNA (rRNA) genes and a control region. The length of the mitogenomes, excluding the control region, ranges from 14,807 bp to 14,977 bp and their GC content range from 21.1% to 26.3%. The cumulated length of the nuclear rRNA genes ranges from 5859 to 5950 bp and the GC content ranged from 40.7% to 43.6%. All sequences were deposited in GenBank (accessions: KX355991-KX356042).

The mean number of sequence reads obtained for each sample was about 34 million (17 million paired-reads). On average 0.3% of the reads were used to assemble the mitogenomes for a mean base coverage of 666 (ranging from 239 to 1,617; SD=416). The mean base coverage of the nuclear ribosomal genes was of 490 (ranging from 99 to 1,083; SD=305). The coverage ratio between mitogenomes and nuclear rRNAs ranged from 0.44 to 7.07 for a mean of 2.23 (SD=2.14). As an illustration, the sequencing coverage of the mitochondrial genome and nuclear ribosomal cluster of *Bichromomyia flaviscutellata* is represented in Fig. 2. Detailed coverage statistics for all species can be found in the Supplementary material.

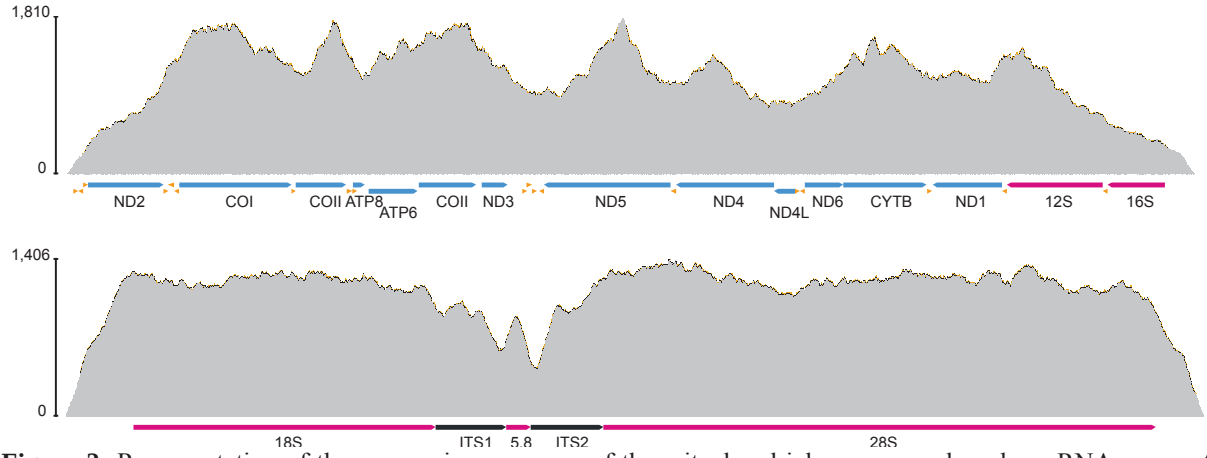


Figure 2: Representation of the sequencing coverage of the mitochondrial genome and nuclear rRNA segment of *Bichromomyia flaviscutellata.*

Phylogenetic analysis

The best partition scheme identified by PartitionFinder is presented in the supplementary material. Overall, the phylogenetic trees obtained exhibit very similar topologies among datasets and inference method. ML trees are presented in Fig. 3 with bootstrap support values and posterior probabilities when the node was also present in the BI tree.

Brumptomyia travassosi was unambiguously retrieved as sister to the remaining New World sand fly species. *Sciopemyia fluviatilis* was placed as sister to all remaining species of *Lutzomyia sensu* Young & Duncan. Other species of the subtribe Lutzomyiina [*Evandromyia* (*Eva*.) *infraspinosa, Pressatia choti* and *Trichopygomyia trichopyga*] were monophyletic with strong statistical support, but their internal relationships differed between mitochondrial and nuclear datasets, and remained poorly supported in the combined dataset. All species of the subtribe Psychodopygina [*Psathyromyia* (*For.*) *aragaoi*, *Viannamyia furcata*, *Bi. flaviscutellata*, *Psychodopygus s. maripaensis*, *Ny. umbratilis*, *Trichophoromyia ininii*] were monophyletic. Their internal relationships also differed between mitochondrial and nuclear datasets; but the reconstruction was strongly supported with the combined dataset. The species of the genus *Micropygomyia* [subtribe Sergentomyiina, *Mi. (Mic.) chassigneti, Mi.* (*Mic*.) *cayennensis* and *Mi.* (*Sau*.) *trinidadensis*] were retrieved as paraphyletic at the base of Psychodopygina.

DISCUSSION

Trees topology and implications for systematics

Galati (1995, 2014) provided the first cladistic analysis of Phlebotominae based on 88 morphological characters (schematised in Fig. 1). Some of our results are consistent with Galati's proposals. The recovering of *Pa.* (*For*.) aragaoi, *Vi. furcata*, *Bi. flaviscutellata, Ny. umbratilis, Th. ininii* and *Ps. maripaensis* into a single clade supports the creation of the subtribe Psychodopygina. Moreover, our analysis corroborates Galati's removal of *Bi. flaviscutellata* from the subgenus *Nyssomyia sensu* Young & Duncan. Conversely, the internal topology of Psychodopygina condradicted Galati's hypothesis by placing *Ps.* (*For.*) *aragaoi* with *Vi. furcata,* and *Bi. flaviscutellata* with *Ps. s. maripaensis* in separate clades. On the contrary, it confirmed the placement of *Nyssomyia* and *Trichophoromyia* as sister groups at a terminal position.

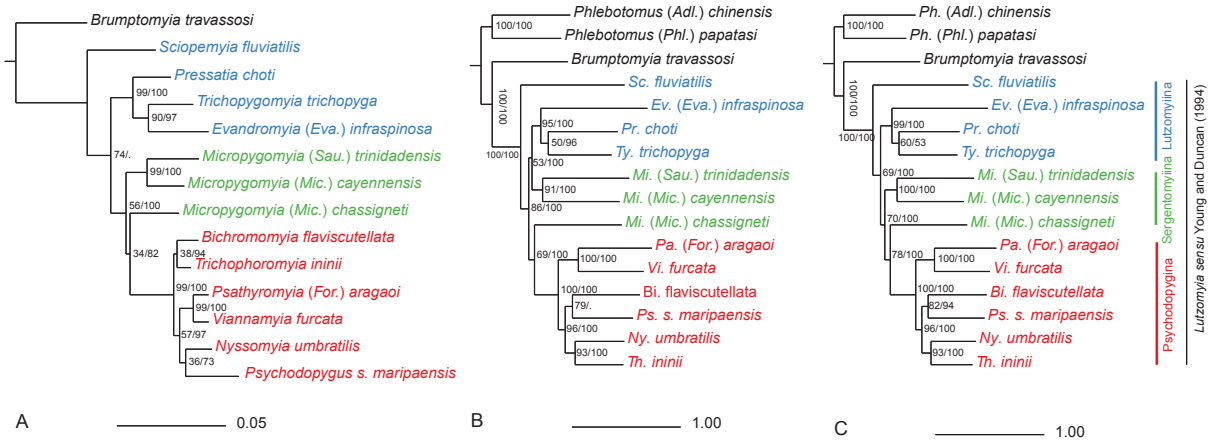


Figure 3: Maximum likelihood phylogenetic tree inferred from **(A)** complete nuclear ribosomal RNA genes (18S and 28S), **(B)** complete mitochondrial protein-coding genes and ribosomal RNA genes, **(C)** the combined datasets. Bootstrap support values expressed in percentage are indicated on the nodes. When the node was also present in the tree inferred by Bayesian analysis, posterior probabilities expressed in percentage are also indicated (on the right).

The subtribe Lutzomyiina is represented by four species in our dataset. *Sc. fluviatilis* was placed as sister species to all others *Lutzomyia sensu* Young & Duncan. This suggests the paraphyly of Lutzomyiina, but Galati (1995) already mentioned that the inclusion of *Sciopemyia* in Lutzomyiina was uncertain. The three other species [*Ev.* (*Eva.*) *infraspinosa, Pr. choti* and *Ty. trichopyga*] were retrieved in a single clade, but their internal relationships differ between the mitochondrial and nuclear datasets. The latter provides a more supported hypothesis, also found in Galati's results, in which *Ev.* (*Eva.*) *infraspinosa* and *Ty. trichopyga* form a sister group at the terminal position in the clade.

The subtribe Sergentomyiina is represented by three species of the genus *Micropygomyia*, which do not form a monophyletic group. *Mi.* (*Mic.*) *cayennensis* and *Mi.* (*Sau.*) *trinidadensis* are recovered as a clade, but the latter does not include *Mi. (Mic.) chassigneti*, which was recovered at the base of Psychodopygina. However, these species are closely related in our tree and the statistical support of their relative positions are too weak to affirm the paraphyly of *Micropygomyia*. On the other hand, their position inside *Lutzomyia sensu* Young & Duncan is strongly supported, which likely refutes the paraphyly of the latter. This has important implications since this taxonomic designation is still largely employed nowadays.

Overall, our results corroborate Galati's classification, which provides new molecular evidence for its use. Inconsistencies were nevertheless observed, but the inclusion of more species in the analysis would be necessary to resolve intergeneric relationships with more confidence. In particular, the addition of Old World *Sergentomyia* specimens in future studies would help concluding on the monophyly of *Lutzomyia sensu* Young & Duncan.

Usefulness of genome skimming for phylogenomics.

Recently, complete mitogenomes have been increasingly used for phylogenetics in a wide range of animals (Comte *et al.*, 2014; Yang *et al.*, 2015; Gibb *et al.*, 2016), a strategy that has been proven successful for providing robust hypotheses. However, the use of a single genomic region for phylogenetic inference may be problematic. Indeed, the results will be susceptible to potential site-specific bias due to biological events such as incomplete lineage sorting and mitochondrial introgression, or analytic issues related to particular evolution rate patterns (Ballard & Whitlock, 2004; McGuire *et al.*, 2007; Som, 2015). For instance, Kopp & True (2002) showed that single-gene datasets of *D. melanogaster* species group produce several strongly supported conflicting clades. Lin & Danforth (2004) studied the differences in patterns of nucleotide substitution between nuclear and mitochondrial DNA and concluded that phylogenetic studies of insects should increasingly focus on nuclear data. The advantage of the genome skimming approach employed in the present study is that it allows recovering the full mitochondrial genome as well as complete nuclear regions in the same workflow. A few topological incongruences were found between the tree based on the nuclear and mitochondrial datasets. However, the conflicting topologies were poorly supported (bootstrap support < 60%) in at least one of the two trees, and may be due to differences in nucleotide substitution rates, rather than truly divergent evolutionary history. As emphasized by our results (see Fig. 3), mitochondrial DNA evolves much faster than nuclear DNA. One or the other genomic region may therefore contain saturated or insufficient phylogenetic signal respectively, depending on the deepness of the relationships inferred. For instance, the internal nodes within Psychodopygina seem better resolved with the mitochondrial dataset, whereas it is the opposite for Lutzomyiina, in which the internal nodes seem deeper.

Implications for metabarcoding and metagenomics

In a recent paper, (Kocher *et al.*, 2016b) have studied the composition of bulk samples of sand flies (corresponding to trap contents) through metabarcoding. This approach consists of extracting the DNA from pools of specimens, amplifying a specific marker allowing to discriminate species, and sequencing those PCR products on an NGS platform. A critical step is the choice of the DNA marker and corresponding PCR primers. In particular, any interspecific variation at the primer-binding site can result in strong amplification bias or detection failure when amplifying a mixture of DNA (Taberlet *et al.*, 2012; Deagle *et al.*, 2014). In other words, a species having more mismatches with the primer than others present in the pool might artificially appear as absent. We previously used the primer pair Ins16S_1,

which was designed for insect metabarcoding (Clarke *et al.*, 2014), but were not able to a priori assess their reliability for the specific case of sand flies. We here show that there are mismatches between the Insect metabarcoding primers and their annealing sites on the 16S gene, but that they are perfectly conserved throughout phlebotominae (at least within the species represented here; Fig. 4). This suggests that the Ins16_1 primers will not produce amplification bias when used for metabarcoding of sand flies. On the contrary, a significant level of variability is found at the binding sites of COI Folmer's primers (Folmer *et al.*, 1994). Therefore, even though the latter have proved to be useful for DNA barcoding of sand flies (Contreras Gutiérrez *et al.*, 2014; Romero-Ricardo *et al.*, 2016), they are less suitable for metabarcoding.

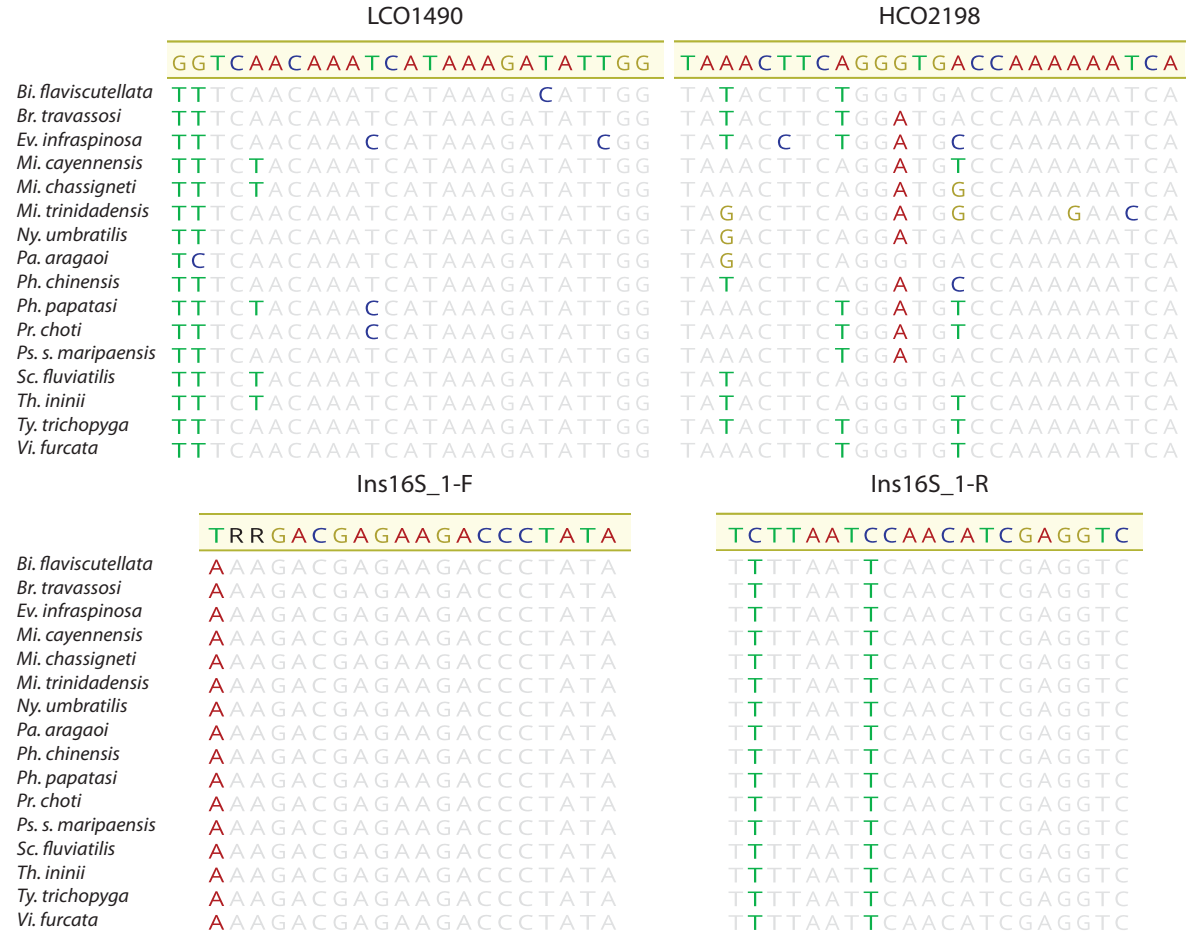


Figure 4: Variability at the binding sites of Folmer's and Ins16S_1 PCR primers among the 16 sand fly species. The sequence of the primers is indicated in the first line of each block. For each species, mismatching positions are colored by nucleotide.

An important point raised by our metabarcoding study is the fact that the PCR step introduces many errors and chimeras that are difficult to detect (Kocher *et al.*, 2016b). An alternative strategy would be to analyse the content of sand fly traps directly through metagenomics (Papadopoulou, Taberlet, & Zinger, 2015), without any PCR procedure. This

"metagenome skimming" has already been demonstrated to be a promising alternative to metabarcoding (Linard *et al.*, 2015). In this context, the mitogenomes obtained through individual-based genome skimming (as in the present study) may be used as a reference database onto which metagenomic reads can be mapped to identify species (Crampton-Platt *et al.*, 2016).

CONCLUDING REMARKS

Our study demonstrates the usefulness of genome skimming for phylogenetic analyses of Phlebotominae. The present work represents one of few molecular studies to assess the phylogenetic relationships of New World sand flies, and to challenge the current taxonomic classifications based on morphological characters. We argue that the application of our strategy to a denser taxonomic sampling can provide great opportunities for studying the evolutionary history of these medically important insects.

ACKNOWLEDGEMENTS:

This work was supported by *Investissement d'Avenir* grants managed by the French *Agence Nationale de la Recherche* (CEBA: ANR-10-LABX-25-01; TULIP: ANR-10-LABX-41, ANR-11-IDEX-0002-02). The collection of sand flies was performed in the framework of the project IAEL (ANR-06-SEST-20). We would like to thank Romuald Carinci, Jean Issaly and Virginie Rougeron for their involvement in the collection and processing of specimens, as well as Sébastien Cally, Jade Bruxaux, Kévin Cilleros and Pierre Solbes for support with bioinformatics. Computations were performed on the EDB lab Cluster which uses a software developed by the Rocks(r) Cluster Group (San Diego Supercomputer Center, University of California, San Diego and its contributors).

REFERENCES

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic acids research* 38: 7–13.

Akhoundi M, Kuhls K, Cannet A, Votýpka J, Marty P, Delaunay P, Sereno D. 2016. A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Neglected Tropical Diseases* 10: e0004349.

Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, Boer M, the WHO Leishmaniasis Control Team. 2012. Leishmaniasis Worldwide and Global Estimates of Its Incidence. *PLoS ONE* 7: e35671.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Molecular Ecology* 13: 729–744.

Beati L, Cáceres AG, Lee JA, Munstermann LE. 2004. Systematic relationships among *Lutzomyia* sand flies (Diptera: Psychodidae) of Peru and Colombia based on the analysis of 12S and 28S ribosomal DNA sequences. *International journal for parasitology* 34: 225–234.

Cameron SL, Lo N, Bourguignon T, Svenson GJ, Evans TA. 2012. A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): Robust support for interfamilial relationships and molecular synapomorphies define major clades. *Molecular Phylogenetics and Evolution* 65: 163–173.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* 17: 540–552.

Clarke LJ, Soubrier J, Weyrich LS, Cooper A. 2014. Environmental metabarcodes for insects: *in silico* PCR reveals potential for taxonomic bias. *Molecular Ecology Resources* 14: 1160–1170.

Cohnstaedt LW, Beati L, Caceres AG, Ferro C, Munstermann LE. 2011. Phylogenetics of the phlebotomine sand fly group Verrucarum (Diptera: Psychodidae: *Lutzomyia*). *The American Journal of Tropical Medicine and Hygiene* 84: 913–922.

Comte L, Murienne J, Grenouillet G. 2014. Species traits and phylogenetic conservatism of climate-induced range shifts in stream fishes. *Nature communications* 5: 5023.

Contreras Gutiérrez MA, Vivero RJ, Vélez ID, Porter CH, Uribe S. 2014. DNA barcoding for the identification of sand fly species (Diptera, Psychodidae, Phlebotominae) in Colombia. *PLoS ONE* 9: e85496.

Crampton-Platt A, Yu DW, Zhou X, Vogler AP. 2016. Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience* 5: 15.

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters* 10: 20140562.

Depaquit J, Grandadam M, Fouque F, Andry PE, Peyrefitte C. 2010. Arthropod-borne viruses transmitted by Phlebotomine sandflies in Europe: a review. *Euro Surveill* 15: 19507.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* 3: 294–299.

Galati EAB. 1995. Phylogenetic systematics of Phlebotominae (Diptera, Psychodidae) with emphasis on American groups. *Bol Dir Malariol San Amb* 35: 133–142.

Galati EAB. 2014. Classificação, morfologia, terminologia e identificação de Adultos: Bioecologia e Identificação de Phlebotominae. In: Lainson R, Rangel EF eds. *Flebotomíneos do Brasil*. Río de Janeiro, 367.

Gibb GC, Condamine FL, Kuch M, Enk J, Moraes-Barros N, Superina M, Poinar HN, Delsuc F. 2016. Shotgun Mitogenomics Provides a Reference Phylogenetic Framework and Timescale for Living Xenarthrans. *Molecular Biology and Evolution* 33: 621–642.

Grace-Lema DM, Yared S, Quitadamo A, Janies DA, Wheeler WC, Balkew M, Hailu A, Warburg A, Clouse RM. 2015. A phylogeny of sand flies (Diptera: Psychodidae: Phlebotominae), using recent Ethiopian collections and a broad selection of publicly available DNA sequence data. *Systematic Entomology* 40: 733–744.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: the beginning of incongruence? *Trends in Genetics* 22: 225–231.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30: 772–780.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, others. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Kim MJ, Kang AR, Jeong HC, Kim KG, Kim I. 2011. Reconstructing intraordinal relationships in Lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). *Molecular Phylogenetics and Evolution* 61: 436–445.

Kocher A, Gantier JC, Holota H, Jeziorski C, Coissac E, Bañuls AL, Girod R, Gaborit P, Murienne J. 2016a. Complete mitochondrial genome of *Lutzomyia* (*Nyssomyia*) *umbratilis* (Diptera: Psychodidae), the main vector of *Leishmania guyanensis*. *Mitochondrial DNA* 27: 4219–4221.

Kocher A, Gantier JC, Gaborit P, Zinger L, Holota H, Valiere S, Dusfour I, Girod R, Bañuls AL, Murienne J. 2016a. Vector soup: high-throughput identification of neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources* doi: 10.1111/1755-0998.12556.

Kocher A, Guilbert É, Lhuillier É, Murienne J. 2015. Sequencing of the mitochondrial genome of the avocado lace bug *Pseudacysta perseae* (Heteroptera, Tingidae) using a genome skimming approach. *Comptes Rendus Biologies* 338: 149–160.

Kocher A, Kamilari M, Lhuillier E, Coissac E, Péneau J, Chave J, Murienne J. 2014. Shotgun assembly of the assassin bug *Brontostoma colossus* mitochondrial genome (Heteroptera, Reduviidae). *Gene* 552: 184–194.

Kopp A, True JR. 2002. Phylogeny of the Oriental *Drosophila melanogaster* species group: a multilocus reconstruction. *Systematic biology* 51: 786–805.

Kück P, Meusemann K. 2010. FASconCAT: Convenient handling of data matrices. *Molecular phylogenetics and evolution* 56: 1115–1118.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution* 29: 1695–1701.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.

Lewis D, Young DG, Fairchild GB, Minter DM. 1977. Proposals for a stable classification of the phlebotomine sandflies (Diptera: Psychodidae). *Systematic Entomology* 2: 319–332.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Lin CP, Danforth BN. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution* 30: 686–702.

Linard B, Crampton-Platt A, Gillett CP, Timmermans MJ, Vogler AP. 2015. Metagenome skimming of insect specimen pools: potential for comparative genomics. *Genome biology and evolution* 7: 1474–1489.

Lins RMMA, Oliveira SG, Souza NA, De Queiroz RG, Justiniano SCB, Ward RD, Kyriacou CP, Peixoto AA. 2002. Molecular evolution of the cacophony IVS6 region in sandflies. *Insect Molecular Biology* 11: 117–122.

Mariette J, Escudié F, Allias N, Salin G, Noirot C, Thomas S, Klopp C. 2012. NG6: Integrated next generation sequencing storage and processing environment. *BMC genomics* 13: 462.

Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L. 2013. Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Medical and Veterinary Entomology* 27: 123–147.

Mazzoni CJ, Gomes CA, Souza NA, Queiroz RG, Justiniano SCB, Ward RD, Kyriacou CP, Peixoto AA. 2002. Molecular evolution of the period gene in sandflies. *Journal of molecular evolution* 55: 553–562.

McGuire JA, Linkem CW, Koo MS, Hutchison DW, Lappin AK, Orange DI, Lemos-Espinal J, Riddle BR, Jaeger JR. 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution* 61: 2879–2897.

Papadopoulou A, Taberlet P, Zinger L. 2015. Metagenome skimming for phylogenetic community ecology: a new era in biodiversity research. *Molecular ecology* 24: 3515–3517.

Ready PD. 2013. Biology of Phlebotomine Sand Flies as Vectors of Disease Agents. *Annual Review of Entomology* 58: 227–250.

Romero-Ricardo L, Lastre-Meza N, Pérez-Doria A, Bejarano EE. 2016. DNA barcoding to identify species of phlebotomine sand fly (Diptera: Psychodidae) in the mixed leishmaniasis focus of the Colombian Caribbean. *Acta Tropica* 159: 125–131.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* 61: 539–542.

Som A. 2015. Causes, consequences and solutions of phylogenetic incongruence. *Briefings in Bioinformatics* 16: 536–548.

Stamatakis A. 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology* 57: 758–771.

Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology* 21: 2045–2050.

Yang X, Cameron SL, Lees DC, Xue D, Han H. 2015. A mitochondrial genome phylogeny of owlet moths (Lepidoptera: Noctuoidea), and examination of the utility of mitochondrial genomes for lepidopteran phylogenetics. *Molecular phylogenetics and evolution* 85: 230–237.

Ye F, Liu T, King SD, You P. 2015. Mitochondrial genomes of two phlebotomine sand flies, *Phlebotomus chinensis* and *Phlebotomus papatasi* (Diptera: Nematocera), the first representatives from the family Psychodidae. *Parasites & vectors* 8: 1–13.

Young DG, Duncan MA. 1994. Guide to the identification and geographic distribution of Lutzomyia sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). *Memoirs of the American Entomological Institute* 54: 1–881.

APPENDICES

Appendix 1: List of the samples used for shotgun sequencing along with GenBank accession numbers

Species (sensu Galati 1995/Young and Duncan 1994)	Voucher code	Country	Collection date	Sex	Genbank accession
Br. travassosi	2698C	French Guiana	22/04/08	M	KX355993, KX356006, KX356019, KX356030
	2699C	French Guiana	22/04/08	M	
	2700C	French Guiana	22/04/08	M	
Bi. flaviscutellata / *L. (Ny.) flaviscutellata*	08MJ	French Guiana	15/11/11	M	KX355999, KX356012, KX356025, KX356036
	34MJ	French Guiana	15/11/11	F	
	48MJ	French Guiana	15/11/11	F	
Ev. (*Eva.*) *infraspinosa* / *L.* (*Ev.*) *infraspinosa*	14MJ	French Guiana	15/11/11	M	KX355996, KX356009, KX356022, KX356033
	1645B	French Guiana	02/10/07	M	
	1666B	French Guiana	02/10/07	M	
Mi. (*Mic.*) *cayennensis* / *L. (Mi.) cayennensis*	09MG	Marie Galante	19/11/12	M	KX356002, KX356015, KX356028, KX356032
	17MG	Marie Galante	19/11/12	F	
	22MG	Marie Galante	19/11/12	M	
Mi. (*Mic.*) *chassignet* / *L. (Pi.) chassignetti i*	1854B	French Guiana	02/10/07	M	KX356001, KX356014, KX356027, KX356041
	1855B	French Guiana	02/10/07	M	
	149B	French Guiana	12/10/07	F	
Mi. (*Sau.*) *trinidadensis* / *L. (Os.) trinidadensis*	03MJ	French Guiana	15/11/11	F	KX355991, KX356004, KX356017, KX356037
	11MJ	French Guiana	15/11/11	M	
Pa. (*For.*) *aragaoi* / *L.* (*Ar.*) *aragaoi*	20TT	Trinidad and Tobago	01/05/12	M	KX356003, KX356016, KX356029, KX356042
	34TT	Trinidad and Tobago	01/05/12	M	
	23CC	French Guiana	06/11/11	F	
Pr. choti/L. (Pr.) choti	03IR	French Guiana	01/06/12	M	KX356000, KX356013, KX356026, KX356038
	05IR	French Guiana	01/06/12	F	
	16MJ	French Guiana	15/11/11	M	
Ps. s. maripaensis / *L. (Ps.) s. maripaensis*	200B	French Guiana	12/10/07	F	KX355994, KX356007, KX356020, KX356031
	205B	French Guiana	12/10/07	F	
	217B	French Guiana	12/10/07	F	
Sc. fluviatilis / *L. (Sc.) fluviatilis*	1660B	French Guiana	02/10/07	M	KX355995, KX356008, KX356021, KX356039
	120B	French Guiana	12/10/07	F	
	1671B	French Guiana	02/10/07	M	
Th. ininii /L. (Th.) ininii	1634B	French Guiana	02/10/07	M	KX355998, KX356011, KX356024, KX356040
	1635B	French Guiana	02/10/07	M	
	1081B	French Guiana	11/10/07	M	
Ty. trichopyga / *L. (Ty.) trichopyga*	151B	French Guiana	12/10/07	F	KX355992, KX356005, KX356018, KX356035
	240B	French Guiana	12/10/07	F	
	82B	French Guiana	12/10/07	F	
Vi. furcata / *L. (Vi.) furcata*	131B	French Guiana	12/10/07	F	KX355997, KX356010, KX356023, KX356034
	138B	French Guiana	12/10/07	F	
	29B	French Guiana	12/10/07	F	

Appendix 2: Sequencing coverage statistics

Species	Total nb. reads	Nb. reads mito	Mean cov. mito.	Nb. reads nuc. rRNAs	Mean cov. nuc. rRNAs	Cov. ratio (mito./nuc. rRNAs)
Br. travassosi	24 230 364	105 261	702	7 350	99	7.07
Bi. flaviscutellata	33 907 252	102 285	668	54 868	744	0.90
Ev. (*Eva.*) *infraspinosa*	19 746 450	45 159	295	7 781	106	2.78
Mi. (*Mic.*) *cayennensis*	33 177 498	221 784	1 390	43 090	527	2.64
Mi. (*Mic.*) *chassignetti*	29 675 092	35 764	239	33 703	456	0.52
Mi. (*Sau.*) *trinidadensis*	53 639 688	76 847	507	18 831	275	1.85
Pa. (*For.*) *aragaoi*	43 843 780	91 476	600	62 995	853	0.70
Pr. choti	40 164 660	248 495	1 617	19 365	279	5.79
Ps. s. maripaensis	32 968 256	124 285	805	77 270	1 083	0.74
Sc. fluviatilis	28 779 048	46 013	309	51 308	703	0.44
Th. ininii	45 519 590	44 290	292	45 385	607	0.48
Ty. trichopyga	27 894 286	83 191	541	10 603	148	3.66
Vi. furcata	33 161 634	104 597	697	34 761	487	1.43
Mean	36 882 353	107 674	700	39 731	542	1.83

Appendix 3: Partitioning scheme of the combined mitochondrial and nuclear dataset (conservative approach) according to PartitionFinder.

Partition	Best model
ATP8_pos1,ND2_pos1,ND3_pos1,ND6_pos1	GTR+I+G
ATP8_pos2,ND2_pos2,ND3_pos2,ND6_pos2	GTR+I+G
ATP6_pos3,ATP8_pos3,COX1_pos3,COX2_pos3,COX3_pos3,CYTB_pos3,ND2_pos3,ND3_pos3,ND6_pos3	GTR+I+G
ATP6_pos1,COX1_pos1,COX2_pos1,COX3_pos1,CYTB_pos1	GTR+G
ATP6_pos2,COX1_pos2,COX2_pos2,COX3_pos2,CYTB_pos2	GTR+I+G
ND1_pos1,ND4L_pos1,ND4_pos1,ND5_pos1	GTR+I+G
ND1_pos2,ND4L_pos2,ND4_pos2,ND5_pos2	GTR+I+G
ND1_pos3,ND4L_pos3,ND4_pos3,ND5_pos3	GTR+G
12S,16S	GTR+I+G
18S,28S	GTR+I+G

CHAPITRE 2: LES COMMUNAUTÉS DE PHLÉBOTOMES FACE AUX MODIFICATIONS DE LA FAUNE DE VERTÉBRÉS

RÉSUMÉ

La compréhension des conséquences des changements environnementaux sur la circulation des maladies vectorielles nécessite une bonne connaissance écologique des arthropodes qui les transmettent. Dans un contexte de déclin de la biodiversité, l'effet de l'altération des communautés d'hôtes vertébrés sur l'abondance des vecteurs hématophages peut grandement conditionner l'existence d'un effet de dilution. **Dans ce chapitre**, nous étudions l'effet des perturbations de la faune de vertébrés sur les communautés de phlébotomes.

L'identification des espèces de phlébotomes se base traditionnellement sur la reconnaissance de caractères morphologiques internes, et nécessite de longues étapes de préparation ainsi que des connaissances taxonomiques pointues. Ceci peut constituer une limite à la réalisation d'études écologiques, surtout dans les régions contenant une grande diversité d'espèces. Dans un **premier article**, nous évaluons l'efficacité du *metabarcoding* pour la caractérisation des communautés de phlébotomes. Nous constituons une base de référence moléculaire pour 40 espèces trouvées en Guyane. Sur la base d'assemblages artificiellement constitués, nous montrons que le *metabarcoding* permet l'identification d'un grand nombre d'espèces dans un même échantillon, même lorsque certaines présentent une abondance relative très faible. Nous appliquons ensuite la méthode à des échantillons de terrain, et montrons qu'elle permet la caractérisation de communautés complexes avec un travail de laboratoire et un coût considérablement réduits par rapport aux approches classiques.

La sensibilité des arthropodes hématophages aux modifications de la faune de vertébrés peut varier selon certaines caractéristiques fonctionnelles. Les phlébotomes disposent d'une capacité de déplacement modeste, et certaines espèces affichent une préférence trophique marquée. Il est donc probable que les communautés de phlébotomes soient impactées par une diminution de l'abondance ou de la diversité des hôtes vertébrés dans le milieu. Aucune étude n'a toutefois réellement été menée en ce sens. **Dans la deuxième partie**, nous présentons les résultats d'une étude écologique préliminaire. Nous employons le metabarcoding pour caractériser les communautés de phlébotomes sur des sites forestiers où la faune de vertébrés a été plus ou moins perturbée dans les régions d'Iracoubo et de Saint-Georges de l'Oyapock. Dans les deux localités étudiées, les résultats indiquent une abondance supérieure et une diversité plus faible sur les sites perturbés, du fait de la prolifération d'une ou deux espèces dominantes. Ces espèces ne sont toutefois pas connues pour être d'importants vecteurs de

leishmaniose. Malgré la congruence des patrons observés, davantage d'efforts sont nécessaires pour savoir si ceux-ci sont réellement liés à une modification de la faune de vertébrés, et s'ils s'accompagnent de conséquences épidémiologiques prédictibles. Ce premier travail exploratoire permet de formuler des hypothèses et d'ouvrir la voie à des études incluant un plus grand nombre d'échantillons.

ARTICLE 1:
VECTOR SOUP:
HIGH-THROUGHPUT IDENTIFICATION OF NEOTROPICAL PHLEBOTOMINE SAND FLIES USING METABARCODING

Arthur Kocher[1,2], Jean-Charles Gantier[3], Pascal Gaborit[4], Lucie Zinger[1], Helene Holota[1], Sophie Valiere[5], Isabelle Dusfour[4], Romain Girod[4], Anne-Laure Bañuls[2] and Jerome Murienne[1*].

Published in *Molecular Ecology Resources* (doi: 10.1111/1755-0998.12556)

[*1]CNRS, Université Toulouse III Paul Sabatier, ENFA; UMR5174 EDB (Laboratoire Evolution et Diversité Biologique), Toulouse, France; [2]UMR MIVEGEC (IRD 224 – CNRS 5290 – Université de Montpellier), 911 Avenue Agropolis, F34394 Montpellier, France; [3]Laboratoire des Identifications Fongiques et Entomo-parasitologiques, Mennecy, France; [4]Medical Entomology Unit, Institut Pasteur de la Guyane, 23 Avenue Pasteur, BP 6010, 97306 Cayenne Cedex, French Guiana ; [5]GeT–PlaGe, Genotoul, INRA Auzeville, 31326 Castanet-Tolosan, France.

ABSTRACT

Phlebotomine sand flies are hematophagous dipterans of primary medical importance. They represent the only proven vectors of leishmaniasis worldwide and are involved in the transmission of various other pathogens. Studying the ecology of sand flies is crucial to understand the epidemiology of leishmaniasis and further control this disease. A major limitation in this regard is that traditional morphological-based methods for sand fly species identifications are time-consuming and require taxonomic expertise. DNA metabarcoding holds great promise in overcoming this issue by allowing the identification of multiple species from a single bulk sample. Here, we assessed the reliability of a short insect metabarcode located in the mitochondrial 16S rRNA for the identification of neotropical sand flies, and constructed a reference database for 40 species found in French Guiana. Then, we conducted a metabarcoding experiment on sand fly mixtures of known content and showed that the method allows an accurate identification of specimens in pools. Finally, we applied metabarcoding to field samples caught in a 1 ha forest plot in French Guiana. Besides providing reliable molecular data for species-level assignations of phlebotomine sand flies, our study proves the efficiency of metabarcoding based on the mitochondrial 16S rRNA for studying sand fly diversity from bulk samples. The application of this high-throughput identification procedure to field samples can provide great opportunities for vector monitoring and eco-epidemiological studies.

INTRODUCTION

Sand flies (Diptera, Psychodidae, Phlebotominae) form a subfamily of hematophagous dipterans that currently comprises close to 1,000 species distributed worldwide. They are of high medical importance as they include all the proven vectors of leishmaniasis and are involved in the transmission of other viral and bacterial diseases (Depaquit *et al.* 2010; Ready 2013). Three sand fly genera are present in the New World, but only species of the *Lutzomyia* genus França, 1924 have been described as vectors of leishmaniasis [the classification proposed by Young & Duncan (1994) will be used throughout the text, but see Supporting Information for the alternative classification of Galati (1995, 2014)]. The *Lutzomyia* genus currently includes 477 described species, among which 56 are proven or suspected vectors of the disease (Killick-Kendrick 1990; Maroli *et al.* 2013).

Investigating the ecology of sand flies is essential to better understand the epidemiology of Leishmaniasis and further control this disease, particularly in regions undergoing important environmental changes (Desjeux 2001; Rotureau 2006; Ready 2008; de Oliveira Miranda *et al.* 2015). A major difficulty in this regard lies in the identification of sand fly specimens, especially in areas such as the Neotropics that harbour a high diversity of species. Traditional morphological identifications of these minute insects are mainly based on digestive and reproductive organs (Young & Duncan 1994; Munstermann 2004; El-Hossary 2006) and represent meticulous and time-consuming tasks (dissection, mounting, examination), even for taxonomic experts.

Molecular methods based on various genetic markers have been developed to overcome this issue (Depaquit 2014). So far, two techniques have been mainly used for molecular identifications of sand flies: the polymerase chain reaction-restriction fragment-length polymorphism (PCR-RFLP; Aransay *et al.* 1999; Terayama *et al.* 2008; Latrofa *et al.* 2012; Tiwary *et al.* 2012; Minter *et al.* 2013) and DNA barcoding (Krüger *et al.* 2011; Kumar *et al.* 2012; Contreras Gutiérrez *et al.* 2014; Al-Ajmi *et al.* 2015; Nzelu *et al.* 2015; Maia *et al.* 2015; Romero-Ricardo *et al.* 2016). DNA barcoding consists in the sequencing of a short standard DNA marker for the taxonomic assignment of a sample by looking for the closest match in a reference database (Hebert *et al.* 2003a). In the last decade, it has emerged as the main alternative for species identification in a wide range of taxa, including medically important arthropods (Ondrejicka *et al.* 2014). More recently, an extended version of DNA barcoding, referred to as DNA metabarcoding, has enabled the identification of multiple species from

bulk samples by mass-PCR-amplification and high-throughput sequencing (Taberlet et al. 2012). This method has been successfully applied to arthropod samples (Yu *et al.* 2012; Ji *et al.* 2013; Gibson *et al.* 2014), opening new avenues for the study of insect biodiversity.

Here, the reliability of DNA metabarcoding to identify sand flies fly species contained in bulk samples is assessed. The current standard barcode for animals is a *c*. 650-bp portion of the mitochondrial cytochrome c oxidase subunit 1 gene (COI) which was already proved useful for the identification of neotropical sand flies (Contreras Gutiérrez *et al.* 2014; Nzelu *et al.* 2015; Romero-Ricardo *et al.* 2016). The standard COI barcode is however not well suited for metabarcoding applications. First, it is too long to be sequenced on the current high-throughput sequencing platforms such as the Illumina Miseq. Moreover, it is virtually impossible to find perfectly conserved primer binding sites in this coding gene, because of high mutation rate at the third codon position (Deagle *et al*. 2014). This may not be a problem when barcoding single specimens, because a few primer-template mismatches will not impede PCR amplification. On the contrary, small variations in primer-template mismatches can lead to significant amplification bias or detection failure when amplifying mixtures of DNA for metabarcoding (Taberlet 2012). Recently, a short fragment of the mitochondrial 16S ribosomal RNA gene was shown to have good properties for insect metabarcoding (Ins16S_1; Clarke et al. 2014). We therefore evaluated the reliability of the Ins16_1 insect metabarcode for the identification of sand flies and constituted the first DNA reference library for species found in French Guiana. We then conducted a metabarcoding assay on artificially constituted sand fly mixtures to assess the accuracy of the method. Finally, we tested the metabarcoding approach on pools of specimens freshly caught in the field.

MATERIAL AND METHOD

Reference library of DNA metabarcodes

Specimens of sand flies were collected with Center for Disease Control (CDC) light traps in various locations in French Guiana between 2007 and 2011 (see Supporting Information for details). Specimens of the same species collected in Trinidad and Tobago in 2012 were also included to assess the validity of the marker over wide geographic ranges. When possible, specimens from both sexes were included in each species, in order to provide molecular evidence for the current acceptations of male-female pairings based on morphological criteria. The samples were processed at Institut Pasteur de la Guyane (Cayenne, French Guiana) and stored at the LIFE laboratory (Mennecy, France). After

dissection, thoraxes were kept for molecular study while heads, wings and genitalia were used for morphological identifications as follows: the pieces were treated according to Abonnenc (1972) and mounted in Euparal medium (Réf 7356.1 Carl Roth GmbH). The morphological and morphometrical data was obtained with the DM5500B LeicaTM microscope associated with the calibration software Leica Application Suite (Version LAS 3.8.0.). Species identifications where done using the taxonomic guide of Young & Duncan (1994), which includes all the references for the descriptions of sand fly species found in French Guiana and Trinidad and Tobago, except for *L. depaquiti* described later (Gantier *et al.* 2006). Specimens were then stored as vouchers. A complete checklist of the sand fly species found in French Guiana is available in the Supporting Information.

DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen. Valencia. CA. USA) and amplified using the Ins16S_1 primer pair (Clarke et al. 2014; Ins16S_1-F. 5'- TRRGACGAGAAGACCCTATA-3'; Ins16S_1-R. 5'- TCTTAATCCAACATCGAGGTC-3'). PCR amplification was performed in 25 µL mixtures containing 2 µL of DNA template, 0.2 µL of AmpliTaq Gold® (5U/ µL; Applied Biosystems, Foster City, CA, USA), 2.5 µL 10X PCR buffer (provided with Amplitaq Gold®), 0.5 µL dNTPs (2.5 mM each, Promega, Madison, WI, USA), 1 µL of each primer (10 µM), 0.25 bovine serum albumin (10mg/mL, Promega), 2.5 µL MgCl2 (25 mM, Applied Biosystems) and nuclease-free water (Promega). Thermocycling conditions consisted in a first denaturation step at 95°C (10 min) followed by 40 cycles of 30s at 95°C, 30s at 50°C and 30s at 72°C, and a final elongation step at 72°C (10 min). To enable the sequencing of multiple PCR products in a single sequencing run, different combinations of tags added to the 5' end of each primer were used (Binladen *et al.* 2007). Tags were eight base pairs long with at least five differences between them to minimize ambiguities in the downstream analyses.

PCR products were pooled and sent to the GeT-PlaGe core facilities of GenoToul (Toulouse, France) for paired-end sequencing on a MiSeq platform (Illumina, San Diego, CA, USA). Prior to sequencing, DNA was purified using the HighPrep PCR system (MAGBIO GENOMICS, Gaithersburg, MD, USA) and used for library construction with the Illumina NEXTflex PCR-Free DNA sequencing kit following the supplier's instructions (Bioo Scientific corp., Austin, TX, USA). Sequencing data was pre-processed and stored on the NG6 platform (Mariette *et al.* 2012) and all downstream bioinformatics analyses were performed on computing facilities of the GenoToul bioinformatics platform (Toulouse, France).

Sequence reads were analysed using the OBITools package (Boyer *et al.* 2016) following Riaz *et al.* (2011). Pair-end reads were aligned and merged, taking into account the phred quality scores for consensus construction and alignment score computation. Reads were then assigned to their corresponding sample, based on the tagged primer sequences. Low quality reads (with alignment scores below 50, containing Ns or shorter than 50 bp) were removed. Identical reads were dereplicated while keeping the coverage information (i.e. number of reads merged). For each sample, only the most abundant sequence was kept. Those that were represented by less than 50 reads were not considered. The resulting DNA reference database was then formatted to be used for taxonomic assignment of sequences. The bash script used for these bioinformatics steps and the DNA reference database in OBITools-extended fasta format are available in the Supporting Information.

The sequences were then aligned with ClustalW2 (Larkin *et al.* 2007) using default parameters. To allow comparisons with other DNA barcoding studies, pairwise genetic distances were computed using the Kimura-2-parameter (K2P) distance model (Kimura 1980). A neighbour-joining tree was generated (Saitou & Nei 1987) to analyse the clustering patterns among species. The R package 'spider' (R Development Core Team 2014, Brown *et al.* 2012) was then used to compute distance statistics and evaluate the reliability of the metabarcode in assigning samples to the correct species using the 'best match' criteria (Meier *et al*. 2006). In order to assess the robustness of the results to various analytical procedures, we performed the same analyses with a different alignment algorithm, muscle (Edgar 2004), or by taking into account the secondary structure of the 16S rRNA using RNAsalsa (Stocsits et al. 2009), as well as using different genetic distance models (JC89 and TN93; Jukes & Cantor 1969; Tamura & Nei 1993).

Metabarcoding validation

Next, a metabarcoding assay was conducted on DNA mixtures of known content, hereafter referred to as "mock communities" (Fig. 1). The first type of mock community consisted of specimens from two distinct species (*L.* (*Trichophoromyia*) *ininii* and *L.* (*Trichopygomyia*) *trichopyga*) that were pooled at ratios of one to 10 or one to 50 prior to bulk DNA extraction. We tested two DNA extraction protocols, adapted from Qiagen extraction kits and Chelex-based protocols (Casquet *et al.* 2012; Ji *et al.* 2013, see Supporting Information for more details). The first one is designed to maximize DNA quality whereas the second is cheap and requires little manipulation. Each extraction protocol was performed

twice for both abundance ratios. Because the concentration of PCR inhibitors and template DNA can influence amplification success and the formation of PCR artefacts (Qiu *et al.* 2001), each resulting DNA solution was diluted 1, 10 and 100 fold in nuclease-free water (Promega) prior to PCR amplifications. The second type of mock community consisted of a pool of DNA from 10 morphologically identified specimens belonging to distinct species. In this case the DNA was extracted individually prior to pooling because the specimens were also included in the reference library.

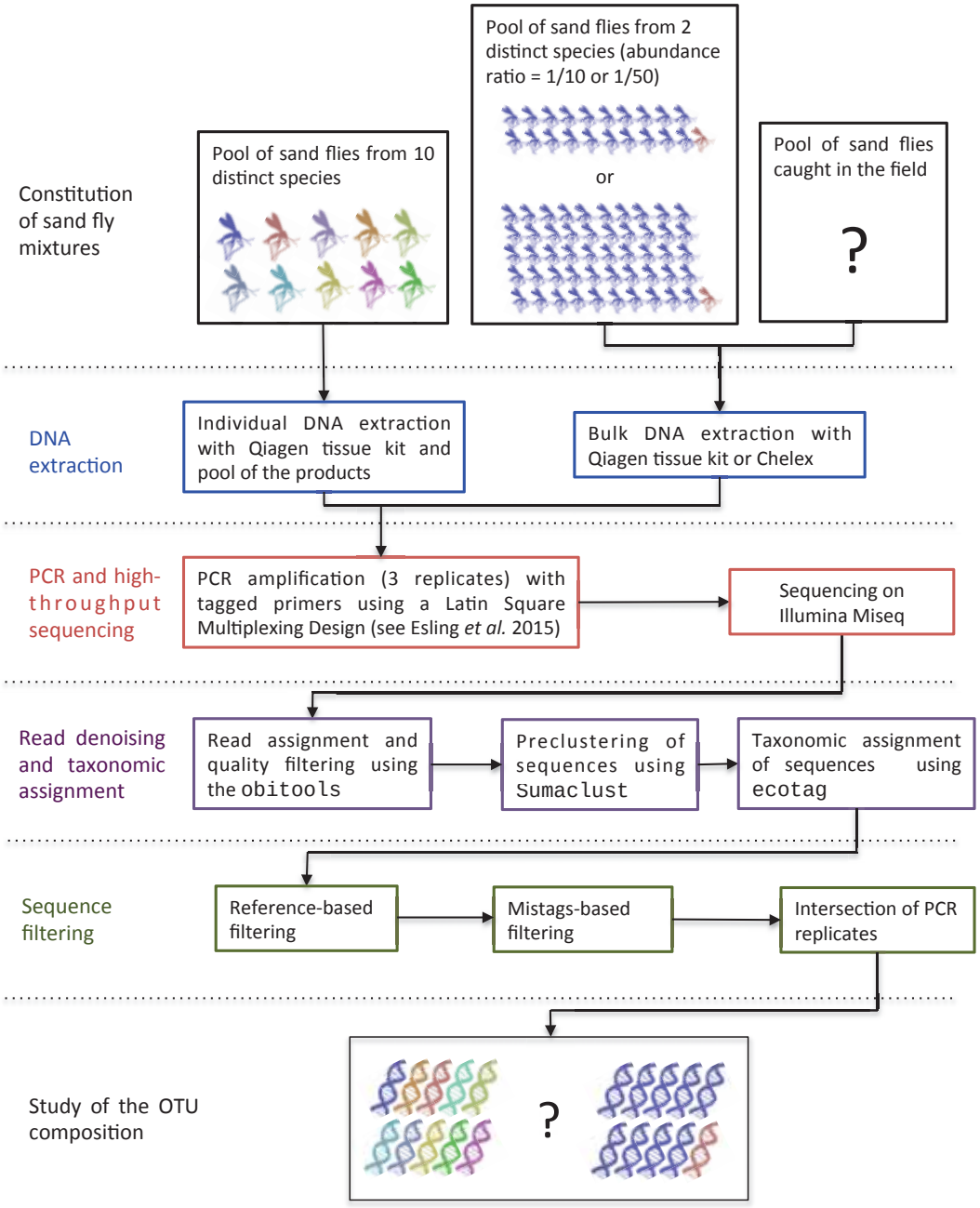


Figure 1: Schematization of the metabarcoding experiment workflow.

Three independent PCR replicates were made for each sample following the protocol described above. For PCR multiplexing, we used the Latin Square Design as described by Esling *et al*. (2015). This design, in which one over two tagged primer combinations are used, has been shown to represent the optimal trade-off between multiplexing capacities and minimizing tag switching events.

Sequencing of amplicons as well as sequence read quality filtering and sample assignment were conducted as described above for the individual specimens. Sequences represented by only one read (singletons) were discarded because they are most likely to arise from artefacts. Sequences were then clustered within each sample using `Sumaclust` (Mercier *et al.* 2013) with a similarity threshold of 99%, as an initial denoising step to remove PCR or sequencing errors (Huse *et al.* 2010). Taxonomic assignment was then performed on our reference DNA database to define operational taxonomic units (OTUs). We further curated the data in order to remove obvious contaminants as well as artefacts such as PCR chimeras, remaining sequencing errors, and missassigned sequence due to tag-switching events (Esling *et al.* 2015). The filtering process was implemented in R and consisted of three steps: (i) a reference based filtering where we kept only the OTUs that had at least 97% identity with its best match in the reference database (ii) filtering of mistagging errors based on OTUs assigned to non-used tag combos (non-critical mistags, see Esling *et al.* 2015 for details) (iii) removal of sequences that were not recovered in at least 2 PCR replicates. The results were then compared to the genuine content of samples to identify false negative and false positive OTUs. For bulk-extracted samples, read abundance ratios were computed to assess whether they were consistent with the specimen ratios. The scripts used for these steps, as well as the sequencing data, are available in the Supporting Information so that the analyses can be reproduced.

Ultimately, metabarcoding was applied to unidentified specimens freshly caught in the field. Sand flies were collected using ten CDC light traps over three nights in a 1 ha forest plot in the region of Iracoubo in French Guyana in October 2015 (WGS84 coordinates: lat=5.25156, long=-53.32059). After each night, the sand flies caught in the traps were killed by freezing, counted and put in 2 mL microcentrifuge tubes (one tube per trap night). Bulk-DNA extraction was performed directly in these tubes after ethanol evaporation using the Chelex protocols. DNA was diluted 10 fold in nuclease-free water prior to PCR amplification. The metabarcoding protocol was then applied as previously described. In order to assess the comprehensiveness of the sampling, an incidence-based species accumulation curve (SAC)

was constructed using the 'exact' method as implemented in the R package 'Vegan' (Oksanen *et al.* 2016). The mean Chao2 estimator of species richness was then computed from 100 permutations along the SAC. Assuming that PCR amplification bias should be consistent across samples, the relative read abundance was weighted by the number of specimens collected in each sample and summed to represent the composition of the sand fly fauna in the whole site.

RESULTS

Reference library of DNA metabarcodes

About 80 species of sand flies have been found in French Guiana, most of them belonging to the genus *Lutzomyia* (Leger *et al.* 1977; Young & Duncan 1994; Gantier *et al.* 2006; a complete checklist of French Guiana species is presented in the Supporting Information). Sequences were obtained for 178 morphologically identified specimens representing 38 species of the genus *Lutzomyia* and two species of the genus *Brumptomyia*. These included *L.* (*Nyssomyia*) *umbratilis*, *L.* (*Nyssomyia*) *flaviscutellata* and *L.* (*Psychodopygus*) *s. maripaensis*, the three proven vectors of *Leishmania* species infecting humans locally (Le Pont & Pajot 1980; Pajot *et al*. 1982; Killick-Kendrick 1990; Fouque *et al*. 2007). All the sequences were deposited in GenBank (accessions KU761608-KU761817). On average, 4.45 specimens per species were sequenced, with more than half of the species being represented by at least 3 specimens and 11 species represented by only one. The sequence length varied between 216 and 222 with an average of 218.5, close to the estimated mean length of the marker in insects (216 bp; Clarke *et al.* 2014). The average base composition was 37.4% adenine (A), 4.5% cytosine (C), 11.6% guanine (G) and 46.5% thymine (T) (GC content 14.8%). 47.3% of sites were identical in all sequences and the mean pairwise identity was 83.3%. Mean intraspecific and interspecific distances by species ranged from 0 to 1.7% and from 0.3 to 22.2% respectively. The distributions of intra and interspecific K2P distances slightly overlapped (Fig. 2a). Nevertheless, for each species taken independently, the maximum intraspecific distance was smaller than the minimum interspecific distance, except for *L.* (*Nyssomyia*) *antunesi* and *L.* (*Nyssomyia*) *yulli pajoti* that were not well distinguished (Fig. 2b, Fig. 3). 98% of the specimens would have been correctly identified if assigned to the species of the closest sequence in the database (without accounting for the species represented by only one specimen). In the neighbour-joining tree generated with the 178 sequences, every species formed a monophyletic group except *L.* (*N.*)

antunesi and *L.* (*N.*) *yulli pajoti* (Fig 3.). All male-female pairings were confirmed and species represented by specimens from both French Guiana and Trinidad and Tobago were also grouped unambiguously. Slight differences were found in the topologies of the NJ trees constructed using the different alignment algorithms and genetic distance models. However, the clustering patterns at the species level and the results of the taxonomic assignation test were strictly identical among these methods.

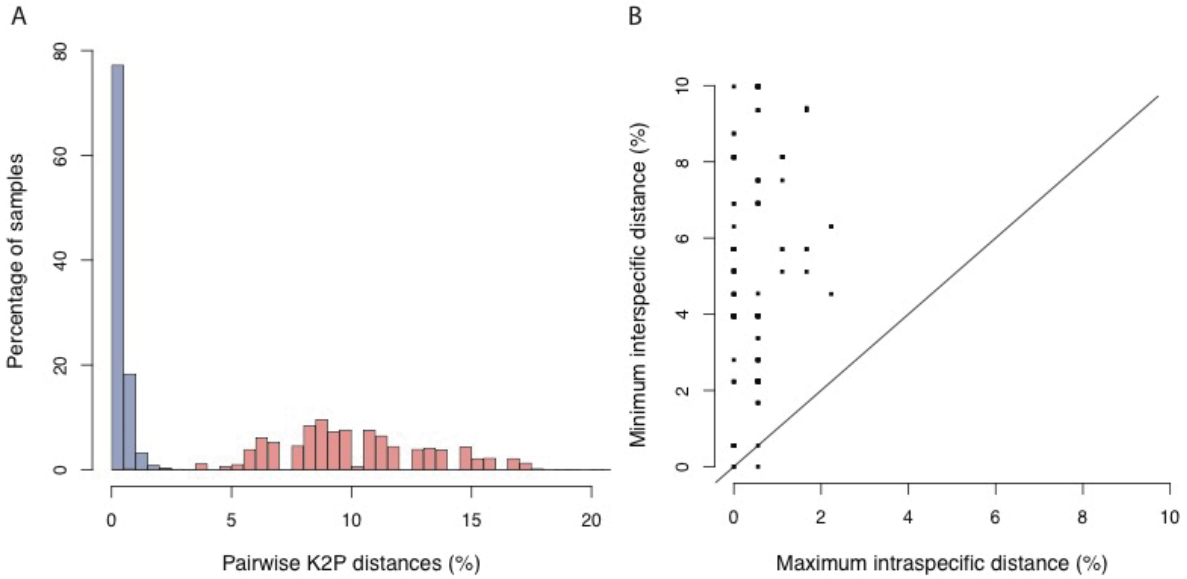


Figure 2: (A) Distribution of intraspecific (in blue) and interspecific (in red) genetic distances among sand fly samples. **(B)** Maximum intraspecific genetic distance against minimum interspecific distance for each sample. The identity line is plotted.

Metabarcoding validation

PCR amplifications were successful for all samples except for three samples extracted with the Chelex protocol that had not been diluted. After sequencing and removal of low quality reads and singletons, 2,626,783 sequencing reads were obtained, representing 59,354 distinct sequences. All expected species were detected in each sample and PCR replicate. Unexpected OTUs were also retrieved but represented only 1.6% of the reads, the majority of which was not identified as any sand fly species. Filtering of the sequences that (i) did not have a close match in the reference database, (ii) resulted from tag switching events and (iii) were not detected in at least 2 PCR replicates, removed 2.6%, 2.4% and 0.5% of the reads respectively. All unexpected OTUs were removed after these steps except in three over 22 bulk-extracted samples where one unexpected species was found. Conversely, all expected species remained detectable in every sample except *L.* (*Evandromyia*) *infraspinosa* in the mixture of ten species and *L.* (*T.*) *ininii* in one of the bulk-extracted sample whatever the DNA dilution prior to PCR. In bulk-extracted samples, read abundance ratios revealed a mean

abundance bias of 2.3 toward *L.* (*T.*) *trichopyga* (σ=1.5). The ratios were fairly consistent among samples and highly similar across PCR replicates. The results of the metabarcoding experiment are synthetized in Table 1.

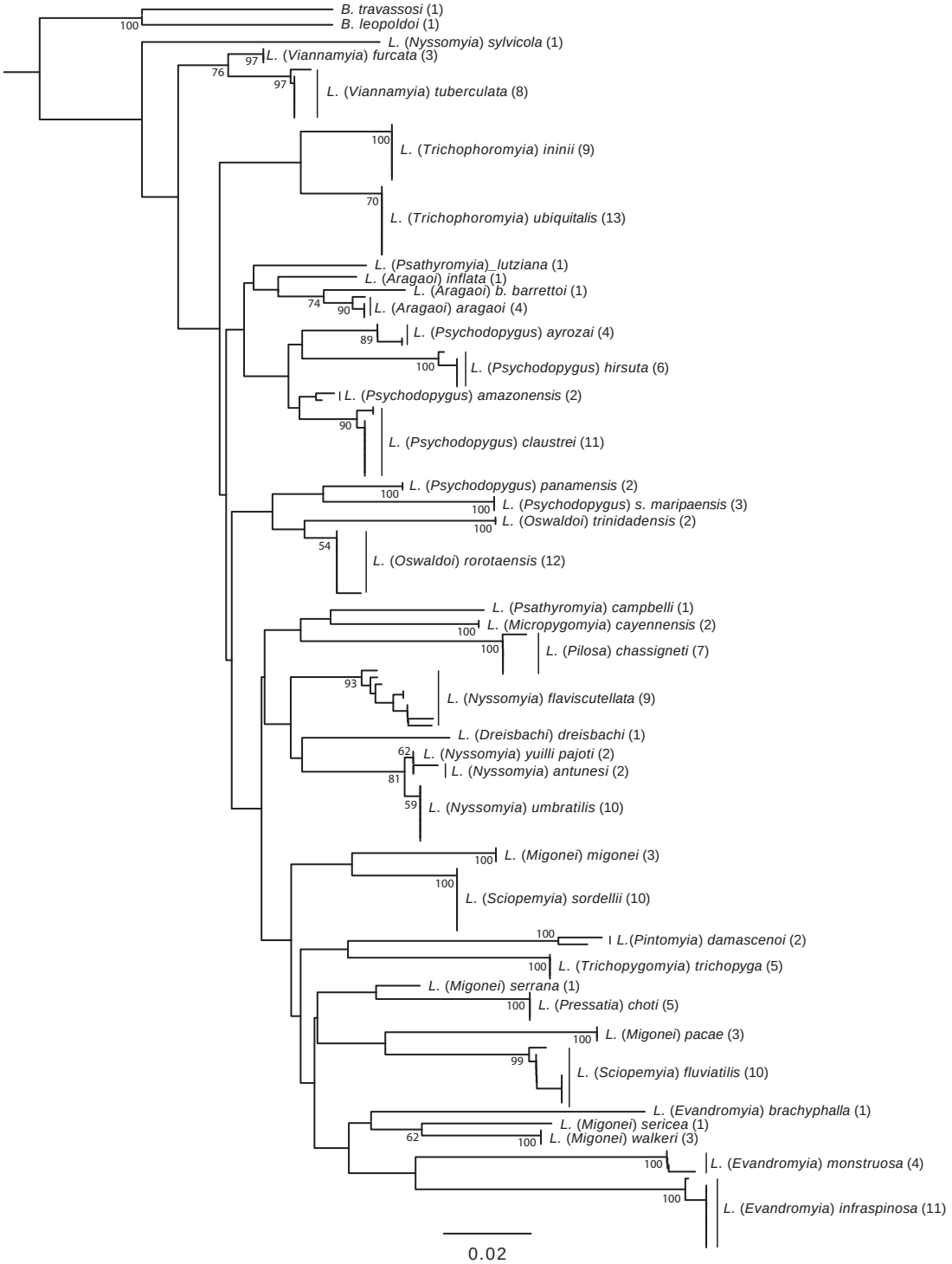


Figure 3: Neighbour-joining tree based on Kimura-2-parameter pairwise distances of the 16S metabarcode from sand fly specimens. Numbers in brackets indicate the number of specimens sequenced per species. Bootstrap support values higher than 50% are indicated on the nodes

.

Table 1: Results of the metabarcoding experiment. Mean numbers of false negative (FN) and false positive (FP) OTUs are presented for each type of sample. For bulk-extracted samples, *L.* (*T.*) *trichopyga* vs *L.* (*T.*) *ininii* read abundance ratios obtained after complete filtering process are presented along with the resulting bias toward *L.* (*T.*) *trichopyga*.

	No filtering		Reference-based filtering		Mistags-based filtering		Intersection of PCR replicates		Reads abundance bias[4]
	FN	FP	FN	FP	FN	FP	FN	FP	
Mix 10[1]	0.0	7.0	0.0	2.0	0.0	0.0	1.0	0.0	NI[5]
Bulk 10[2]	0.0	5.7	0.0	2.0	0.0	0.0	0.0	0.0	2.2
Bulk 50[3]	0.0	6.2	0.0	3.0	0.0	0.6	0.3	0.3	2.5

[1] mixture of DNA extracted from ten distinct sand fly species.
[2] bulk-extracted samples including 2 species with a specimen ratio of 10
[3] bulk-extracted samples including 2 species with a specimen ratio of 50
[4]abundance bias: bias found between the relative abundance of *L.* (*T.*) *trichopyga* and *L.* (*T.*) *ininii* included in the mock mixtures and corresponding read abundance ratio
[5]non-informative

A total of 415 sand flies were collected in 28 trap nights. From these 28 samples, 24 were successfully amplified and sequenced. 25 distinct species were identified. Some sequences closely matched to both *L.* (*N.*) yuilli pajoti and *L.* (*N.*) antunesi and where therefore assigned to the *Nyssomyia* subgenus. The species accumulation curve constructed from the cumulative incidence of species across trap nights reaches an asymptote (Fig. 4). The mean Chao2 estimator oscillates around 25 species along the SAC and converges to a value of 25.08 (SE=0.33). After weighting the read counts by the number of specimens in each sample, the more abundant species was *L.* (*E.*) *infraspinosa* (25.4% of reads). *L.* (*N.*) *umbratilis, L.* (*N.*) *flaviscutellata and L.* (*P.*) *s. maripaensis*, two locally proven vectors of Leishmaniasis, were also found.

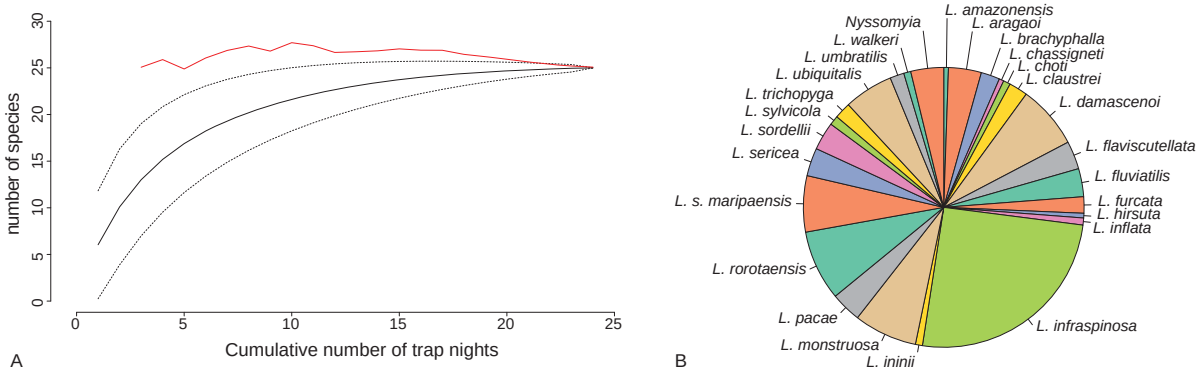


Figure 4: Results of metabarcoding on field samples. **(A)** Species accumulation curve (black solid line) based on the incidence of species trap nights constructed with the 'exact' method along with its confidence interval (black dotted lines). The solid red line indicates the mean Chao2 estimator computed from 100 permutations. **(B)** Overall OTU composition after weighting the read counts by the number of specimens in each sample.

DISCUSSION

DNA metabarcoding represents a promising tool to overcome taxonomic impediments in assessing the sand fly fauna diversity and distribution. Because COI is not the best option for metabarcoding (Deagles et al. 2014), we assessed the efficiency of an insect metabarcode located in the mitochondrial 16S ribosomal RNA gene to discriminate between species of this medically important group. Since the distributions of intra and interspecific genetic distances slightly overlapped, the "barcoding gap" obtained is not as clear as expected for the standard COI barcoding marker (Hebert *et al.* 2003b). The latter has been claimed to be illusive in dipterans (Meier *et al.* 2006), but several evaluations of the COI marker in phlebotomine sand flies did report a barcoding gap between intra and interspecific distances (Kumar *et al.* 2012; Contreras Gutiérrez *et al.* 2014; Nzelu *et al.* 2015; Romero-Ricardo *et al.* 2016). Nevertheless, the assignation of each sequence to the species of its closest match in the database resulted in almost 98% correct identifications. This is remarkable because the choice of a metabarcode implies a compromise between its versatility and its resolution (taxonomic coverage and specificity; Riaz *et al.* 2011). Except for *L.* (*N.*) *antunesi* and *L.* (*N.*) *yuilli pajoti*, the distance tree obtained with the mt 16S rRNA marker clustered every species studied in monophyletic groups. It also unambiguously grouped conspecific individuals sampled in very distant locations (French Guiana and Trinidad and Tobago). This is noteworthy assuming that intraspecific genetic variation correlates with the geographic scale of the sampling (Bergsten *et al.* 2012). Finally, the marker confirmed the associations of all conspecific males and females, highlighting the reliability of traditional morphological methods in this regard. Most of the subgenera represented in the dataset were not retrieved as monophyletic groups in the NJ tree. It is important to keep in mind that the latter only aims at visualising the clustering of individuals belonging to the same species in order to assess the reliability of the marker for taxonomic assignations based on genetic distances. It should not be confused with a phylogenetic analysis and cannot be used to infer deeper taxonomic relationships. The three sand fly species proven to transmit of Leishmaniasis in French Guiana are all represented in the reference database and well discriminated with the metabarcode. Therefore, the latter is adapted for eco-epidemiological studies in this country. Nevertheless, its use for metabarcoding studies in other regions requires first to generate reference sequences libraries for the species of interest locally and to assess the ability of the marker to discriminate them.

Our metabarcoding experiments showed that it was possible to accurately identify the content of sand fly mixtures even in samples exhibiting a relatively high species diversity or

abundance imbalance. A meticulous data curation is essential to avoid significant overestimations of diversity in metabarcoding studies. It represents the most critical step of analysis pipelines (Kunin *et al.* 2010; Yu *et al.* 2012; Esling *et al.* 2015). Here, the sequence filtering process was very efficient in removing false positive OTUs. Only one unexpected species was detected in three of the 22 bulk-extracted samples. Because the bulk-samples were constituted of entire specimens that were morphologically identified without dissection and mounting, it is difficult to exclude that these false positive OTUs were due to the inclusion of misidentified specimens rather than resulting from molecular biases. These results highlight the advantage of working with a relatively comprehensive reference database. First, it allowed a reference-based filtering that eliminated the majority of unexpected sequences. Second, defining OTUs in a reference-supervised way ensures OTU taxonomic consistency. De novo OTU picking is far from being a trivial matter (Bonder *et al.* 2012; Sun *et al.* 2012). We tested several combinations of preclustering and clustering algorithms with various parameters on our data (Sumaclust, Swarm and MCL, Van Dongen 2000; Mahé *et al.* 2014). All led to important discrepancies between the resulting de novo OTUs and taxonomic assignments of the sequences (data not shown). On the other hand, the obvious disadvantage of such a stringent reference-based analysis pipeline is that it is blind to species that are not represented in the reference database. Nevertheless, the same sequencing data can be reanalysed later on with an updated reference database. Moreover, it is possible to check for the presence of unknown OTUs in the data by relaxing the stringency of the filtering process. However, great care should be taken by using this approach, as it will inevitably increase the number of artifactual OTUs (false positives). We have looked in the field samples for sequences that had between 90% and 97% identity with their best match in the reference database and a read abundance of at least 200 in at least 2 PCR replicates. This way, we identified two OTUs that may correspond to *Lutzomyia* species not represented in the reference database: one having 94.1% identity with *L. hirsuta* and another having 94.6% identity with *L. sylvicola*.

Both extraction protocols led to very satisfying results. Since the Chelex-based protocol is cheap and time efficient, it represents a good candidate for routine use. Nevertheless, it lacks a purification step, which can lead to PCR amplification failure and quick degradation of the DNA in the samples, making them not suitable for long term storage. This likely explains the amplification failures in three out of the four Chelex-extracted samples that were

not diluted prior to PCR. Dilution of the samples in distilled water at a minimum of 1:10 is therefore advisable for samples extracted with the Chelex protocol.

Several metabarcoding studies have shown significant incongruences between the genuine content of mock communities and amplicon relative abundances (Amend *et al.* 2010; Yu *et al.* 2012; Piñol *et al.* 2015). Thus, the conclusion is usually that metabarcoding is inappropriate for quantification purposes. In pool-extracted samples, read abundance ratios tended to overestimate the proportion of *L.* (*T.*) *trichopyga* (the more abundant species in the mixture), with a mean bias of 2.3. This is may be due to an actual difference in biomass, even though *L.* (*T.*) *trichopyga and L.* (*T.*) *ininii* have a relatively close typical size. Variability in primer-template mismatches have been shown to result in amplification bias of up to five orders of magnitude (Bru *et al.* 2008; Piñol *et al.* 2015). However, the complete mitochondrial genome of several sand fly species, including *L.* (*T.*) *ininii* and *L.* (*T.*) *trichopyga*, were sequenced, and all exhibit 100% similarity at the Ins16S_1 primers binding sites (Kocher *et al.* 2015, Ye *et al.* 2015 and further unpublished data). Various others technical and biological factors can impact the quantitative signal and lead to discrepancies (Pompanon *et al.* 2012), but this is beyond the scope of this paper. However, the bias found here was less than one order of magnitude and was fairly consistent across samples and PCR replicates. Relative read abundance may therefore be indicative when studying sand fly mixtures, at least for inter-sample comparisons.

The results obtained on our field samples indicate that metabarcoding was efficient for characterizing the local sand fly community. The species accumulation curve reaches an asymptote and the Chao2 estimator converges to a value extremely close to the observed number of species, giving good confidence in the comprehensiveness of the sampling. In total, 25 distinct species were identified in the plot, which is comparable to the results of other studies conducted in the region (Rotureau *et al.* 2006; Fouque *et al.* 2007). As mentioned above, great caution has to be taken for quantitative interpretation of metabarcoding. Assuming that PCR amplification bias is consistent across samples, the results presented in Fig. 4b can be regarded as a proxy of species composition for between site comparisons, but are not to be confused with real proportions of specimens.

The use of metabarcoding allowed us to obtain these results with little laboratory work (28 samples processed instead of 415 if we had treated each specimen individually). However, the number of sand flies caught here was relatively low. The added value of metabarcoding in terms of cost and time-effectiveness depends on the number of specimens

pooled prior to bulk DNA extraction and on the number of samples multiplexed for high-throughput sequencing. Using the present protocol, we estimate that 10,000 specimens can be analysed by a single person in 20 days and for a total cost of less than 5,000 euros. As a mater of comparison, just mounting those specimens for morphological identification would take more than 200 days (considering 10 min per specimen and 8 hours of work per day).

In conclusion, this study highlight that the insect metabarcode developed by Clarke *et al.* (2014) is reliable for species-level assignations of phlebotomine sand flies and that metabarcoding using this marker can accurately identify the composition of sand fly mixtures. The application of the present protocol to field samples can significantly improve the cost and time effectiveness of data generation on sand fly fauna. Moreover, the primers used here have been estimated to amplify over 90% of insect taxa among the publicly available data at the time of their design and can therefore be employed widely. With adequate reference databases, this approach could be extended to concomitantly analyse other insect vectors communities such as mosquitoes, opening new perspectives for eco-epidemiological studies.

ACKNOWLEDGMENTS

This work was supported by “Investissement d’Avenir” grants managed by Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; TULIP: ANR-10-LABX-41, ANR-11-IDEX-0002-02) as well as project METABAR (ANR-11-BSV7-0020). The collection of sand flies was performed in the framework of the French National Project IAEL: ANR-06- SEST-20. We would like to thank Romuald Carinci, Jean Issaly, Virginie Rougeron and Benoit de Thoisy for their involvement in the collection and processing of the samples, Pierre Taberlet and Eric Coissac for fruitful discussions about metabarcoding and Pieternel Verschuren for her kind revision of the English text.

DATA ACCESSIBILITY

- Sand fly specimens: the specimens used for the DNA reference database are kept as voucher at the LIFE laboratory (Mennecy, France). Pictures can be sent on demand.
- DNA reference database: All sequences were published in GenBank (accessions KU761608-KU761817). The sequences are also available in the OBITools extended fasta format together with the corresponding taxonomic database in the Supporting information (S4).

- Metabarcoding data: sequencing data of mock communities and field samples is available in the Supporting information (S5). We provide the data in fasta format after dereplication and removal of singletons and in tabular format after taxonomic assignment of the sequences.
- Analysis pipeline: bash and R script used for the bioinformatic treatment of the data are available in the Supporting Information (S6 and S7).

AUTHOR CONTRIBUTIONS

A.K. and J.M. designed the study. JC.G., P.G., I.D., R.G. and AL.B. were involved in the collection of samples on the field. JC.G. and P.G. conducted the morphological identifications. A.K., H.H. and S.V. performed the laboratory work. A.K., J.M. and L.Z. participated in the analysis of the data. A.K. wrote the manuscript and all authors contributed in its improvement.

REFERENCES

Abonnenc E (1972) Les phlébotomes de la région éthiopienne (Diptera, Psychodidae). *Mémoires ORSTOM*, **55.**

Al-Ajmi RA, Ayaad TH, Al-Anazi M, Al-Ahdal MN, Al-Qahtani AA (2015) Molecular identification of natural sand fly species populations inferred from ITS2 rRNA gene in Saudi Arabia. *Journal of Environmental Biology*, **36**.

Alvar J, Vélez ID, Bern C *et al.* (2012) Leishmaniasis Worldwide and Global Estimates of Its Incidence. *PLoS ONE*, **7**, e35671.

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular Ecology*, **19**, 5555–5565.

Aransay AM, Scoulica E, Chaniotis B, Tselentis Y (1999) Typing of sandflies from Greece and Cyprus by DNA polymorphism of 18S rRNA gene. *Insect Molecular Biology*, **8**, 179–184.

Bergsten J, Bilton DT, Fujisawa T *et al.* (2012) The Effect of Geographical Scale of Sampling on DNA Barcoding. *Systematic Biology*, sys037.

Binladen J, Gilbert MTP, Bollback JP *et al.* (2007) The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLoS ONE*, **2**.

Bonder MJ, Abeln S, Zaura E, Brandt BW (2012) Comparing clustering and pre-processing in taxonomy analysis. *Bioinformatics*, **28**, 2891–2897.

Boyer F, Mercier C, Bonin A *et al.* (2016) obitools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, **16**, 176-182.

Brown SD, Collins RA, Boyer S *et al.* (2012) Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, **12**, 562–565.

Bru D, Martin-Laurent F, Philippot L (2008) Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. *Applied and environmental microbiology*, **74**, 1660–1663.

Casquet J, Thebaud C, Gillespie RG (2012) Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. *Molecular Ecology Resources*, **12**, 136–141.

Clarke LJ, Soubrier J, Weyrich LS, Cooper A (2014) Environmental metabarcodes for insects: *in silico* PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, **14**, 1160-1170.

Contreras Gutiérrez MA, Vivero RJ, Vélez ID, Porter CH, Uribe S (2014) DNA Barcoding for the Identification of Sand Fly Species (Diptera, Psychodidae, Phlebotominae) in Colombia. *PLoS ONE*, **9**, e85496.

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, **10**, 20140562.

Depaquit J (2014) Molecular systematics applied to Phlebotomine sandflies: Review and perspectives. *Infection, Genetics and Evolution*, **28**, 744–756.

Depaquit J, Grandadam M, Fouque F, Andry PE, Peyrefitte C (2010) Arthropod-borne viruses transmitted by Phlebotomine sandflies in Europe: a review. *Euro Surveill*, **15**, 19507.

Desjeux P (2001) The increase in risk factors for leishmaniasis worldwide. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **95**, 239–243.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792–1797.

El-Hossary S (2006) Morphological characteristics for sand fly taxonomy. *Research and Training Center on Vector of Diseases, Ain Shams University, Ain Shams*, 1–25.

Esling P, Lejzerowicz F, Pawlowski J (2015) Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, **43**, 2513–2524.

Gantier JC, Gaborit P, Rabarison P (2006) Les phlébotomes (Diptera: Psychodidae) de la Guyane Française. Description du mâle de *Lutzomyia* (*Trichopygomyia*) *depaquiti* n. sp. *Parasite*, **13**, 11–15.

Galati, E. A. B. (1995). Phylogenetic systematics of Phlebotominae (Diptera, Psychodidae) with emphasis on American groups. *Bol Dir Malariol San Amb,* **35**(Supl 1), 133-142.

Galati EAB (2014) Phlebotominae (Diptera, Psychodidae) Classificação, Morfologia, Terminologia e Identificação de Adultos. Apostila. Departamento de Epidemiologia Faculdade de Saúde Pública Universidade de São Paulo, Brasil, 133 pp.

Gibson J, Shokralla S, Porter TM *et al.* (2014) Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, **111**, 8007–8012.

Hebert PD, Cywinska A, Ball SL, others (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, Waard JR de (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, S96–S99.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, **12**, 1889–1898.

Ji Y, Ashton L, Pedley SM *et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, **16**, 1245–1257.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. *Mammalian protein metabolism*, **3**, 132.

Killick-Kendrick R (1990) Phlebotomine vectors of the leishmaniases: a review. *Medical and Veterinary Entomology*, **4**, 1–24.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, **16**, 111–120.

Kocher A, Gantier J-C, Holota H *et al.* (2015) Complete mitochondrial genome of *Lutzomyia* (*Nyssomyia*) *umbratilis* (Diptera: Psychodidae), the main vector of *Leishmania guyanensis*. *Mitochondrial DNA*, 1–3.

Krüger A, Strüven L, Post RJ, Faulde M (2011) The sandflies (Diptera: Psychodidae, Phlebotominae) in military camps in northern Afghanistan (2007–2009), as identified by morphology and DNA "barcoding". *Annals of Tropical Medicine & Parasitology*, **105**, 163–176.

Kumar NP, Srinivasan R, Jambulingam P (2012) DNA barcoding for identification of sand flies (Diptera: Psychodidae) in India. *Molecular Ecology Resources*, **12**, 414–420.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology*, **12**, 118–123.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Latrofa MS, Annoscia G, Dantas-Torres F, Traversa D, Otranto D (2012) Towards a rapid molecular identification of the common phlebotomine sand flies in the Mediterranean region. *Veterinary parasitology*, **184**, 267–270.

Leger N, Abonnenc E, Pajot F-X, Kramer R, Claustre J (1977) Liste commentée des phlébotomes de la Guyane Française. *Cahiers ORSTOM. Série Entomologie Médicale et Parasitologie*, **15**, 217–232.

Lightburn E, Meynard JB, Morand J et al. (2001) Surveillance épidémiologique des leishmanioses tégumentaires en Guyane Synthèse de données militaires sur 10 ans. *Méd Trop*, **62**, 545–553.

Mahé F, Rognes T, Quince C, De Vargas C, Dunthorn M (2014) Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ*, **2**, e593.

Maia C, Parreirab R, Cristóvão JM, Afonso MO, Campino L (2015) Exploring the utility of phylogenetic analysis of cytochrome oxidase gene subunit I as a complementary tool to classical taxonomical identification of phlebotomine sand fly species (Diptera, Psychodidae) from southern Europe. *Acta Tropica*, **144**, 1–8.

Mariette J, Escudié F, Allias N *et al.* (2012) NG6: Integrated next generation sequencing storage and processing environment. *BMC genomics*, **13**, 462.

Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L (2013) Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Medical and Veterinary Entomology*, **27**, 123–147.

Meier R, Shiyang K, Vaidya G, Ng PK (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic biology*, **55**, 715–728.

Mercier C, Boyer F, Bonin A, Coissac E (2013) SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences. In: *Programs and Abstracts of the SeqBio 2013 workshop. Abstract*, pp. 27–29.

Minter LM, Yu T, Florin DA *et al.* (2013) Molecular Identification of Sand Flies (Diptera: Psychodidae) in Eastern North America by Using PCR-RFLP. *Journal of Medical Entomology*, **50**, 920–924.

Munstermann LE (2004) Phlebotomine sand flies, the Psychodidae. *Biology of disease vectors*, 141–151.

Nzelu CO, Cáceres AG, Arrunátegui-Jiménez MJ *et al.* (2015) DNA barcoding for identification of sand fly species (Diptera: Psychodidae) from leishmaniasis-endemic areas of Peru. *Acta tropica*, **145**, 45–51.

de Oliveira Miranda DE, da Silva Sales KG, da Gloria Faustino MA *et al.* (2015) Ecology of sand flies in a low-density residential rural area, with mixed forest/agricultural exploitation, in north-eastern Brazil. *Acta tropica*, **146**, 89–94.

Oksanen J, Blanchet FG, Kindt R *et al.* (2016). vegan: Community Ecology Package. R package version 2.3-3.

Ondrejicka DA, Locke SA, Morey K, Borisenko AV, Hanner RH (2014) Status and prospects of DNA barcoding in medically important parasites and vectors. *Trends in Parasitology*, **30**, 582–591.

Peçanha Brazil R, Fuzzari Rodrigues AA, Andrade Filho JD (2015) Sand Fly Vectors of *Leishmania* in the Americas - A Mini Review. *Entomology, Ornithology & Herpetology: Current Research*, **4**.

Piñol J, Mir G, Gomez-Polo P, Agustí N (2015) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, **15**, 819–830.

Pompanon F, Deagle BE, Symondson WOC *et al.* (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, **21**, 1931–1950.

Qiu X, Wu L, Huang H *et al.* (2001) Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology*, **67**, 880–887.

R Development Core Team (2014) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2012. ISBN 3-900051-07-0.

Ready PD (2008) Leishmaniasis emergence and climate change. *Revue Scientifique Et Technique-Office International Des Epizooties*, **27**.

Ready PD (2013) Biology of Phlebotomine Sand Flies as Vectors of Disease Agents. *Annual Review of Entomology*, **58**, 227–250.

Riaz T, Shehzad W, Viari A *et al.* (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, e145–e145.

Romero-Ricardo L, Lastre-Meza N, Pérez-Doria A, Bejarano EE (2016) DNA barcoding to identify species of phlebotomine sand fly (Diptera: Psychodidae) in the mixed leishmaniasis focus of the Colombian Caribbean. *Acta Tropica*, **159**, 125–131.

Rotureau B, Gaborit P, Issaly J et al. (2006) Diversity and ecology of sand flies (Diptera: Psychodidae: Phlebotominae) in coastal French Guiana. *American Journal of Tropical Medicine and Hygiene*, **75**, 62.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, **4**, 406–425.

Stocsits RR, Letsch H, Hertel J, Misof B, Stadler PF (2009) Accurate and efficient reconstruction of deep phylogenies from structured RNAs. *Nucleic Acids Research*, **37**(18), 6184-6193.

Sun Y, Cai Y, Huse SM *et al.* (2012) A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Briefings in Bioinformatics*, **13**, 107–121.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, **21**, 2045–2050.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, **10**, 512–526.

Terayama Y, Kato H, Gomez EA *et al.* (2008) Molecular Typing of Sand Fly Species (Diptera, Psychodidae, Phlebotominae) from Areas Endemic for Leishmaniasis in Ecuador by PCR-RFLP of 18S Ribosomal RNA Gene. *Journal of Veterinary Medical Science*, **70**, 907–913.

Tiwary P, Kumar D, Rai M, Sundar S (2012) PCR-RFLP Based Method for Molecular Differentiation of Sand Fly Species *Phlebotomus argentipes*, *Phlebotomus papatasi* and *Sergentomyia babu* Found in India. *Journal of Medical Entomology*, **49**, 1515–1518.

Van Dongen S (2000) A cluster algorithm for graphs. *Report-Information systems*, 1–40.

Ye F, Liu T, King SD, You, P (2015). Mitochondrial genomes of two phlebotomine sand flies, *Phlebotomus chinensis* and *Phlebotomus papatasi* (Diptera: Nematocera), the first representatives from the family Psychodidae. *Parasites & vectors*, ***8*(1)**, 1-13.

Young DG, Duncan MA (1994) *Guide to the identification and geographic distribution of Lutzomyia sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae)*. Gainesville, FL: Associated Publishers.

Yu DW, Ji Y, Emerson BC *et al.* (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.

APPENDICES

Appendix 1: Checklist of sand fly species reported in French Guiana. The classifications of Young & Duncan (1994) and Galati (1995, 2014) are presented.

Young & Duncan, 1994		Galati, 1995, 2005			
Genus	**Subgenus (Group)**	**Genus**	**Subgenus**	**Species**	**Authors**
Brumptomyia		*Brumptomyia*		*pintoi*	(Costa Lima, 1932)
				travassosi	(Mangabeira Fo, 1942)
				spinosipes	(Floch & Abonnenc, 1943)
Lutzomyia	*Lutzomyia*	*Lutzomyia*	*Tricholateralis*	*carvalhoi*	(Damasceno, Causey & Arouck,1945)
				gomezi	(Nitzulescu, 1931)
				spathotrichia	Martins, Falcao & Da Silva, 1963
	Psathyromyia	*Psathyromyia*	*Psathyromyia*	*abonnenci*	(Floch & Chassignet, 1947)
				campbelli	(Damasceno, Causey & Arouck,1945)
				dendrophylla	(Mangabeira Fo, 1942)
				punctigeniculata	(Floch & Abonnenc, 1944)
				scaffi	(Damasceno & Arouck, 1956)
				shannoni	(Dyar, 1929)
			Forattiniella	*lutziana*	(Costa Lima, 1932)
	Aragaoi (gp.)			*aragaoi*	(Costa Lima, 1932)
				b. barrettoi	(Mangabeira Fo, 1942)
				castilloi	León, Mollinedo, Le Pont, 2009*
				brasiliensis	(Costa Lima, 1932)
				inflata	(Floch & Abonnenc, 1944)
	Dreisbachi (gp)	*Psathyromyia*	*Xiphomyia*	*dreisbachi*	(Causey & Damasceno, 1945)
	Evandromyia	*Evandromyia*	*Evandromyia*	*brachyphalla*	(Mangabeira Fo, 1941)
				infraspinosa	(Mangabeira Fo, 1941)
				pinottii	(Damasceno & Arouck, 1956)
			Barrettomyia	*monstruosa*	(Floch & Abonnenc, 1944)
	Saulensis (gp)		*Evandromyia*	*saulensis*	(Floch & Abonnenc, 1944)
	Baityi	*Migonemyia*	*Blancasmyia*	*moucheti*	Pajot & le Pont, 1978
	Ungrouped			*bursiformis*	(Floch & Chassignet, 1944)
	Micropygomyia	*Micropygomyia*	*Micropygomyia*	*c. cayennensis*	(Floch & Abonnenc, 1941)
				micropyga	(Mangabeira Fo, 1942)
	Pilosa			*chassigneti*	(Floch & Abonnenc, 1944)
				pilosa	(Damasceno & Causey, 1944)
	Oswaldoi		*Sauromyia*	*peresi*	(Mangabaira Fo, 1942)
				pusilla	Dias, Martins, Falcao & Silva, 1986
				rorotaensis	(Floch & Abonnenc, 1944)
				trinidadensis	(Newstead, 1922)
	Micropygomyia	*Martinsmyia*		*quadrispinosa*	(Floch & Chassignet, 1947)
	Migonei (gp)	*Migonemyia*	*Migonemyia*	*migonei*	(França, 1920)
		Pintomyia	*Pifanomyia*	*pacae*	(Floch & Abonnenc, 1943)
		Evandromyia	*Aldamyia*	*sericea*	Floch & Abonnenc, 1944
				sp of Baduel	(Floch & Abonnenc, 1945)
				walkeri	(Newstead, 1914)
				williamsi	(Damasceno, Causey & Arouck,1945)

Young & Duncan, 1994		Galati, 1995, 2005			
Genus	**Subgenus (Group)**	**Genus**	**Subgenus**	**Species**	**Authors**
	Nyssomyia	*Nyssomyia*		*anduzei*	(Rozeboom, 1942)
				antunesi	(Coutinho, 1939)
				bibinae	Léger & Abonnenc, 1988
				sylvicola	(Floch & Abonnenc, 1945)
				umbratilis	Ward & Frahia, 1977
				whitmani	(Artunes & Coutinho, 1939)
				yuilli pajoti	Abonnenc, Léger & Fauran, 1979
		Bichromomyia		*flaviscutellata*	(Mangabaira Fo, 1942)
	Pintomyia	*Pintomyia*	*Pintomyia*	*damascenoi*	(Mangabeira Fo, 1941)
	Pressatia	*Pressatia*		*choti*	(Floch & Abonnenc, 1941)
				equatorialis	(Mangabeira Fo, 1942)
				triacantha	(Mangabeira Fo, 1942)
				trispinosa	(Mangabeira Fo, 1942)
	Psychodopygus	*Psychodopygus*		*amazonensis*	(Root, 1934)
				ayrozai	(Barretto & Coutinho, 1940)
				bispinosa	(Fairchild & Hertig, 1951)
				claustrei	Abonnenc, Léger & Fauran, 1979
				corossoniensis	le Pont & Pajot, 1978
				davisi	(Root, 1934)
				dorlinsis	le Pont & Desjeux, 1982
				geniculata	(Mangabeira Fo, 1941)
				guyanensis	(Floch & Abonenc, 1941)
				h. hirsuta	(Mangabeira Fo, 1942)
				nocticola	Young, 1973
				panamensis	(Shannon, 1926)
				paraensis	(Costa Lima, 1941)
				squamiventris.	(Floch & Abonnenc, 1946)
	Sciopemyia	*Sciopemyia*		*fluviatilis*	((Floch & Abonnenc, 1944)
				sordellii	(Shannon & Del Ponte, 1927)
	Trichophoromyia	*Trichophoromyia*		*brachipyga*	(Mangabeira Fo, 1942)
				flochi	(Abonnenc & Chassignet, 1948)
				ininii	(Floch & Abonnenc, 1943)
				ubiquitalis	(Mangabaira Fo, 1942
	Trichopygomyia	*Trichopygomyia*		*depaquiti*	Gantier, Gaborit & Rabarison, 2006
				longispina	(Mangabaira Fo, 1942)
				trichopyga	(Floch & Abonnenc, 1945)
	Verrucarum (gp)	*Pintomyia*	*Pifanomyia*	*odax*	(Fairchild & Hertig, 1961)
				serrana	(Damasceno & Arouck, 1949)
	Viannamyia	*Viannamyia*		*fariasi*	(Damasceno, Causey & Arouck,1945)
				furcata	(Mangabeira Fo, 1941)
				tuberculata	(Mangabeira Fo, 1941)
Warileya		*Warileya*		*fourgassiensis*	Le Pont & Desjeux, 1984

*In "Phlébotomes d'Équateur et de Bolivie: descriptions de Psathyromyia castilloi n. sp. et Psychodopygus luisleoni n. sp.", Bulletin de la Société entomologique de France, 2009, 114 (1): 83-89, the authors state that a specimen from the Ininii river in French Guiana, firstly identified as L. b. barrettoi by Abonnenc and conserved at the NMNH in Paris with the voucher code 2418C has been identified as L. castilloi after revision.

Appendix 2: List of the samples used for the metabarcode reference database along with the GenBank accession numbers of the corresponding sequences

Voucher	Species	Sex	Country	Locality	Coll. date	Genbank AN
50TT	*B. leopoldoi*	M	Trinidad and Tobago	Chaguaramas	05/01/12	KU761817
211MJ	*B. travassosi*	M	French Guiana	Montjoly	15/11/11	KU761697
23CC	*L.* (*Aragaoi*) *aragaoi*	F	French Guiana	Cacao	11/06/11	KU761704
2614B	*L.* (*Aragaoi*)*aragaoi*	F	French Guiana	Saint-Georges	02/10/07	KU761711
20TT	*L.* (*Aragaoi*) *aragaoi*	M	Trinidad and Tobago	Chaguaramas	05/01/12	KU761696
34TT	*L.* (*Aragaoi*) *aragaoi*	M	Trinidad and Tobago	Chaguaramas	05/01/12	KU761817
1938B	*L.* (*Aragaoi*) *b. barrettoi*	NA	French Guiana	Saint-Georges	10/02/07	KU761689
2882B	*L.* (*Aragaoi*) *inflata*	F	French Guiana	Saint-Georges	02/10/07	KU761801
LD1	*L.* (*Dreisbachi*) *dreisbachi*	NA	French Guiana	Saint-Georges	02/10/07	KU761817
1856B	*L.* (*Evandromyia*) *brachyphalla*	M	French Guiana	Saint-Georges	10/02/07	KU761683
243B	*L.* (*Evandromyia*) *infraspinosa*	F	French Guiana	Cacao	10/12/07	KU761706
45B	*L.* (*Evandromyia*) *infraspinosa*	F	French Guiana	Cacao	10/12/07	KU761817
47B	*L.* (*Evandromyia*) *infraspinosa*	F	French Guiana	Cacao	10/12/07	KU761817
42MJ	*L.* (*Evandromyia*) *infraspinosa*	F	French Guiana	Montjoly	15/11/11	KU761817
1197A	*L.* (*Evandromyia*) *infraspinosa*	F	French Guiana	Saint-Georges	19/04/07	KU761629
1202A	*L.* (*Evandromyia*) *infraspinosa*	F	French Guiana	Saint-Georges	02/10/07	KU761631
14MJ	*L.* (*Evandromyia*) *infraspinosa*	M	French Guiana	Montjoly	15/11/11	KU761650
1939B	*L.* (*Evandromyia*) *infraspinosa*	M	French Guiana	Saint-Georges	02/10/07	KU761689
1645B	*L.* (*Evandromyia*) *infraspinosa*	M	French Guiana	Saint-Georges	10/02/07	KU761653
1665B	*L.* (*Evandromyia*) *infraspinosa*	M	French Guiana	Saint-Georges	10/02/07	KU761658
1666B	*L.* (*Evandromyia*) *infraspinosa*	M	French Guiana	Saint-Georges	10/02/07	KU761659
196B	*L.* (*Evandromyia*) *monstruosa*	F	French Guiana	Cacao	10/12/07	KU761691
2691B	*L.* (*Evandromyia*) *monstruosa*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2720B	*L.* (*Evandromyia*) *monstruosa*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2723B	*L.* (*Evandromyia*) *monstruosa*	F	French Guiana	Saint-Georges	02/10/07	KU761711
04IR	*L.* (*Micropygomyia*) *cayennensis*	F	French Guiana	Ile Royale	06/01/12	KU761612
02IR	*L.* (*Micropygomyia*) *cayennensis*	M	French Guiana	Ile Royale	06/01/12	KU761609
2613B	*L.* (*Migonei*) *migonei*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2630B	*L.* (*Migonei*) *migonei*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2692B	*L.* (*Migonei*) *migonei*	F	French Guiana	Saint-Georges	02/10/07	KU761711
71MJ	*L.* (*Migonei*) *pacae*	F	French Guiana	Montjoly	15/11/11	KU761817
2628B	*L.* (*Migonei*) *pacae*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2997B	*L.* (*Migonei*) *pacae*	F	French Guiana	Saint-Georges	02/10/07	KU761817
2880B	*L.* (*Migonei*) *sericea*	F	French Guiana	Saint-Georges	02/10/07	KU761791
2921B	*L.* (*Migonei*) *serrana*	F	French Guiana	Saint-Georges	02/10/07	KU761817
139B	*L.* (*Migonei*) *walkeri*	F	French Guiana	Cacao	10/12/07	KU761643
2625B	*L.* (*Migonei*) *walkeri*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2726B	*L.* (*Migonei*) *walkeri*	F	French Guiana	Saint-Georges	02/10/07	KU761711
104TT	*L.* (*Nyssomyia*) *antunesi*	M	Trinidad and Tobago	Arima Forest	05/02/12	KU761621
25TT	*L.* (*Nyssomyia*) *antunesi*	M	Trinidad and Tobago	Chaguaramas	05/01/12	KU761711
190B	*L.* (*Nyssomyia*) *flaviscutellata*	F	French Guiana	Cacao	10/12/07	KU761687
36B	*L.* (*Nyssomyia*) *flaviscutellata*	F	French Guiana	Cacao	10/12/07	KU761817
34MJ	*L.* (*Nyssomyia*) *flaviscutellata*	F	French Guiana	Montjoly	15/11/11	KU761817
48MJ	*L.* (*Nyssomyia*) *flaviscutellata*	F	French Guiana	Montjoly	15/11/11	KU761817
2883B	*L.* (*Nyssomyia*) *flaviscutellata*	F	French Guiana	Saint-Georges	02/10/07	KU761801
2890B	*L.* (*Nyssomyia*) *flaviscutellata*	F	French Guiana	Saint-Georges	02/10/07	KU761801
2891B	*L.* (*Nyssomyia*) *flaviscutellata*	F	French Guiana	Saint-Georges	02/10/07	KU761801
2892B	*L.* (*Nyssomyia*) *flaviscutellata*	F	French Guiana	Saint-Georges	02/10/07	KU761801
08MJ	*L.* (*Nyssomyia*) *flaviscutellata*	M	French Guiana	Montjoly	15/11/11	KU761617
25MJ	*L.* (*Nyssomyia*) *sylvicola*	M	French Guiana	Montjoly	15/11/11	KU761710
244B	*L.* (*Nyssomyia*) *umbratilis*	F	French Guiana	Cacao	10/12/07	KU761707
3B	*L.* (*Nyssomyia*) *umbratilis*	F	French Guiana	Cacao	10/11/07	KU761817
25B	*L.* (*Nyssomyia*) *umbratilis*	F	French Guiana	Cacao	10/12/07	KU761708
1897B	*L.* (*Nyssomyia*) *umbratilis*	M	French Guiana	Saint-Georges	02/10/07	KU761686
1912B	*L.* (*Nyssomyia*) *umbratilis*	M	French Guiana	Saint-Georges	02/10/07	KU761687
1925B	*L.* (*Nyssomyia*) *umbratilis*	M	French Guiana	Saint-Georges	02/10/07	KU761688
1926B	*L.* (*Nyssomyia*) *umbratilis*	M	French Guiana	Saint-Georges	02/10/07	KU761688
1927B	*L.* (*Nyssomyia*) *umbratilis*	M	French Guiana	Saint-Georges	02/10/07	KU761688
1669B	*L.* (*Nyssomyia*) *umbratilis*	M	French Guiana	Saint-Georges	10/02/07	KU761661
1688B	*L.* (*Nyssomyia*) *umbratilis*	M	French Guiana	Saint-Georges	10/02/07	KU761671
2B	*L.* (*Nyssomyia*) *yuilli pajoti*	F	French Guiana	Cacao	10/11/07	KU761817
6B	*L.* (*Nyssomyia*) *yuilli pajoti*	F	French Guiana	Cacao	10/11/07	KU761817
123B	*L.* (*Oswaldoi*) *rorotaensis*	F	French Guiana	Cacao	10/12/07	KU761634
136B	*L.* (*Oswaldoi*) *rorotaensis*	F	French Guiana	Cacao	10/12/07	KU761640
2635B	*L.* (*Oswaldoi*) *rorotaensis*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2636B	*L.* (*Oswaldoi*) *rorotaensis*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2637B	*L.* (*Oswaldoi*) *rorotaensis*	F	French Guiana	Saint-Georges	02/10/07	KU761711
1869B	*L.* (*Oswaldoi*) *rorotaensis*	M	French Guiana	Saint-Georges	10/02/07	KU761685
1870B	*L.* (*Oswaldoi*) *rorotaensis*	M	French Guiana	Saint-Georges	10/02/07	KU761686
1670B	*L.* (*Oswaldoi*) *rorotaensis*	M	French Guiana	Saint-Georges	02/10/07	KU761661
1898B	*L.* (*Oswaldoi*) *rorotaensis*	M	French Guiana	Saint-Georges	02/10/07	KU761686
1871B	*L.* (*Oswaldoi*) *rorotaensis*	M	French Guiana	Saint-Georges	02/10/07	KU761686
1674B	*L.* (*Oswaldoi*) *rorotaensis*	M	French Guiana	Saint-Georges	10/02/07	KU761664

1675B	*L.* (*Oswaldoi*) *rorotaensis*	M	French Guiana	Saint-Georges	10/02/07	KU761665
03MJ	*L.* (*Oswaldoi*) *trinidadensis*	F	French Guiana	Montjoly	15/11/11	KU761611
11MJ	*L.* (*Oswaldoi*) *trinidadensis*	M	French Guiana	Montjoly	15/11/11	KU761631
149B	*L.* (*Pilosa*) *chassigneti*	F	French Guiana	Cacao	10/12/07	KU761647
49B	*L.* (*Pilosa*) *chassigneti*	F	French Guiana	Cacao	10/12/07	KU761817
2621B	*L.* (*Pilosa*) *chassigneti*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2876B	*L.* (*Pilosa*) *chassigneti*	F	French Guiana	Saint-Georges	02/10/07	KU761783
2878B	*L.* (*Pilosa*) *chassigneti*	F	French Guiana	Saint-Georges	02/10/07	KU761783
1854B	*L.* (*Pilosa*) *chassigneti*	M	French Guiana	Saint-Georges	10/02/07	KU761681
1855B	*L.* (*Pilosa*) *chassigneti*	M	French Guiana	Saint-Georges	10/02/07	KU761682
2609B	*L.* (*Pintomyia*) *damascenoi*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2961B	*L.* (*Pintomyia*) *damascenoi*	F	French Guiana	Saint-Georges	02/10/07	KU761817
05IR	*L.* (*Pressatia*) *choti*	F	French Guiana	Ile Royale	06/01/12	KU761614
44MJ	*L.* (*Pressatia*) *choti*	F	French Guiana	Montjoly	15/11/11	KU761817
2748B	*L.* (*Pressatia*) *choti*	F	French Guiana	Saint-Georges	02/10/07	KU761711
03IR	*L.* (*Pressatia*) *choti*	M	French Guiana	Ile Royale	06/01/12	KU761610
16MJ	*L.* (*Pressatia*) *choti*	M	French Guiana	Montjoly	15/11/11	KU761672
1853B	*L.* (*Psathyromyia*) *campbelli*	M	French Guiana	Saint-Georges	10/02/07	KU761680
10B	*L.* (*Psathyromyia*) *lutziana*	F	French Guiana	Cacao	10/11/07	KU761624
2618B	*L.* (*Psychodopygus*) *amazonensis*	F	French Guiana	Saint-Georges	02/10/07	KU761711
1677B	*L.* (*Psychodopygus*) *amazonensis*	M	French Guiana	Saint-Georges	10/02/07	KU761666
1650B	*L.* (*Psychodopygus*) *ayrozai*	M	French Guiana	Saint-Georges	10/02/07	KU761655
94TT	*L.* (*Psychodopygus*) *ayrozai*	F	Trinidad and Tobago	Arima Forest	05/02/12	KU761817
38TT	*L.* (*Psychodopygus*) *ayrozai*	F	Trinidad and Tobago	Chaguaramas	05/01/12	KU761817
93TT	*L.* (*Psychodopygus*) *ayrozai*	M	Trinidad and Tobago	Arima Forest	05/02/12	KU761817
4B	*L.* (*Psychodopygus*) *claustrei*	F	French Guiana	Cacao	10/11/07	KU761817
220B	*L.* (*Psychodopygus*) *claustrei*	F	French Guiana	Cacao	10/12/07	KU761700
89B	*L.* (*Psychodopygus*) *claustrei*	F	French Guiana	Cacao	10/12/07	KU761817
2725B	*L.* (*Psychodopygus*) *claustrei*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2729B	*L.* (*Psychodopygus*) *claustrei*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2938B	*L.* (*Psychodopygus*) *claustrei*	F	French Guiana	Saint-Georges	02/10/07	KU761817
1690B	*L.* (*Psychodopygus*) *claustrei*	M	French Guiana	Saint-Georges	02/10/07	KU761671
1863B	*L.* (*Psychodopygus*) *claustrei*	M	French Guiana	Saint-Georges	02/10/07	KU761684
1680B	*L.* (*Psychodopygus*) *claustrei*	M	French Guiana	Saint-Georges	10/02/07	KU761667
1681B	*L.* (*Psychodopygus*) *claustrei*	M	French Guiana	Saint-Georges	10/02/07	KU761668
1687B	*L.* (*Psychodopygus*) *claustrei*	M	French Guiana	Saint-Georges	10/02/07	KU761670
194B	*L.* (*Psychodopygus*) *hirsuta*	F	French Guiana	Cacao	10/12/07	KU761690
198B	*L.* (*Psychodopygus*) *hirsuta*	F	French Guiana	Cacao	10/12/07	KU761692
203B	*L.* (*Psychodopygus*) *hirsuta*	F	French Guiana	Cacao	10/12/07	KU761694
2888B	*L.* (*Psychodopygus*) *hirsuta*	F	French Guiana	Saint-Georges	02/10/07	KU761801
1648B	*L.* (*Psychodopygus*) *hirsuta*	M	French Guiana	Saint-Georges	10/02/07	KU761654
1668B	*L.* (*Psychodopygus*) *hirsuta*	M	French Guiana	Saint-Georges	10/02/07	KU761660
49MJ	*L.* (*Psychodopygus*) *panamensis*	F	French Guiana	Montjoly	15/11/11	KU761817
12MJ	*L.* (*Psychodopygus*) *panamensis*	M	French Guiana	Montjoly	15/11/11	KU761637
200B	*L.* (*Psychodopygus*) *s. maripaensis*	F	French Guiana	Cacao	10/12/07	KU761693
205B	*L.* (*Psychodopygus*) *s. maripaensis*	F	French Guiana	Cacao	10/12/07	KU761695
217B	*L.* (*Psychodopygus*) *s. maripaensis*	F	French Guiana	Cacao	10/12/07	KU761698
120B	*L.* (*Sciopemyia*) *fluviatilis*	F	French Guiana	Cacao	10/12/07	KU761633
2881B	*L.* (*Sciopemyia*) *fluviatilis*	F	French Guiana	Saint-Georges	02/10/07	KU761801
2889B	*L.* (*Sciopemyia*) *fluviatilis*	F	French Guiana	Saint-Georges	02/10/07	KU761801
2900B	*L.* (*Sciopemyia*) *fluviatilis*	F	French Guiana	Saint-Georges	02/10/07	KU761801
2913B	*L.* (*Sciopemyia*) *fluviatilis*	F	French Guiana	Saint-Georges	02/10/07	KU761817
1857B	*L.* (*Sciopemyia*) *fluviatilis*	M	French Guiana	Saint-Georges	10/02/07	KU761684
1706B	*L.* (*Sciopemyia*) *fluviatilis*	M	French Guiana	Saint-Georges	02/10/07	KU761673
1660B	*L.* (*Sciopemyia*) *fluviatilis*	M	French Guiana	Saint-Georges	10/02/07	KU761657
1671B	*L.* (*Sciopemyia*) *fluviatilis*	M	French Guiana	Saint-Georges	10/02/07	KU761662
1705B	*L.* (*Sciopemyia*) *fluviatilis*	M	French Guiana	Saint-Georges	10/02/07	KU761673
191B	*L.* (*Sciopemyia*) *sordellii*	F	French Guiana	Cacao	10/12/07	KU761688
72B	*L.* (*Sciopemyia*) *sordellii*	F	French Guiana	Cacao	10/12/07	KU761817
46B	*L.* (*Sciopemyia*) *sordellii*	F	French Guiana	Cacao	10/12/07	KU761817
2675B	*L.* (*Sciopemyia*) *sordellii*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2897B	*L.* (*Sciopemyia*) *sordellii*	F	French Guiana	Saint-Georges	02/10/07	KU761801
1908B	*L.* (*Sciopemyia*) *sordellii*	M	French Guiana	Saint-Georges	02/10/07	KU761686
1914B	*L.* (*Sciopemyia*) *sordellii*	M	French Guiana	Saint-Georges	02/10/07	KU761687
1672B	*L.* (*Sciopemyia*) *sordellii*	M	French Guiana	Saint-Georges	10/02/07	KU761663
1771B	*L.* (*Sciopemyia*) *sordellii*	M	French Guiana	Saint-Georges	10/02/07	KU761675
1837B	*L.* (*Sciopemyia*) *sordellii*	M	French Guiana	Saint-Georges	10/02/07	KU761678
233B	*L.* (*Trichophoromyia*) *ininii*	F	French Guiana	Cacao	10/12/07	KU761702
11B	*L.* (*Trichophoromyia*) *ininii*	F	French Guiana	Cacao	10/12/07	KU761630
13B	*L.* (*Trichophoromyia*) *ininii*	F	French Guiana	Cacao	10/12/07	KU761644
21B	*L.* (*Trichophoromyia*) *ininii*	F	French Guiana	Cacao	10/12/07	KU761699
1193A	*L.* (*Trichophoromyia*) *ininii*	F	French Guiana	Saint-Georges	02/10/07	KU761629
1636B	*L.* (*Trichophoromyia*) *ininii*	M	French Guiana	Saint-Georges	02/10/07	KU761652
1637B	*L.* (*Trichophoromyia*) *ininii*	M	French Guiana	Saint-Georges	02/10/07	KU761652
1638B	*L.* (*Trichophoromyia*) *ininii*	M	French Guiana	Saint-Georges	02/10/07	KU761652
1683B	*L.* (*Trichophoromyia*) *ininii*	M	French Guiana	Saint-Georges	10/02/07	KU761669
234B	*L.* (*Trichophoromyia*) *ubiquitalis*	F	French Guiana	Cacao	10/12/07	KU761703
117B	*L.* (*Trichophoromyia*) *ubiquitalis*	F	French Guiana	Cacao	10/12/07	KU761629
182B	*L.* (*Trichophoromyia*) *ubiquitalis*	F	French Guiana	Cacao	10/12/07	KU761677
81B	*L.* (*Trichophoromyia*) *ubiquitalis*	F	French Guiana	Cacao	10/12/07	KU761817
2701B	*L.* (*Trichophoromyia*) *ubiquitalis*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2716B	*L.* (*Trichophoromyia*) *ubiquitalis*	F	French Guiana	Saint-Georges	02/10/07	KU761711

2760B	*L.* (*Trichophoromyia*) *ubiquitalis*	F	French Guiana	Saint-Georges	02/10/07	KU761711
1704B	*L.* (*Trichophoromyia*) *ubiquitalis*	M	French Guiana	Saint-Georges	02/10/07	KU761672
1707B	*L.* (*Trichophoromyia*) *ubiquitalis*	M	French Guiana	Saint-Georges	02/10/07	KU761673
1709B	*L.* (*Trichophoromyia*) *ubiquitalis*	M	French Guiana	Saint-Georges	02/10/07	KU761673
1726B	*L.* (*Trichophoromyia*) *ubiquitalis*	M	French Guiana	Saint-Georges	02/10/07	KU761674
1731B	*L.* (*Trichophoromyia*) *ubiquitalis*	M	French Guiana	Saint-Georges	02/10/07	KU761674
1739B	*L.* (*Trichophoromyia*) *ubiquitalis*	M	French Guiana	Saint-Georges	02/10/07	KU761674
240B	*L.* (*Trichopygomyia*) *trichopyga*	F	French Guiana	Cacao	10/12/07	KU761705
25CC	*L.* (*Trichopygomyia*) *trichopyga*	F	French Guiana	Cacao	11/06/11	KU761709
151B	*L.* (*Trichopygomyia*) *trichopyga*	F	French Guiana	Cacao	10/12/07	KU761651
82B	*L.* (*Trichopygomyia*) *trichopyga*	F	French Guiana	Cacao	10/12/07	KU761817
2714B	*L.* (*Trichopygomyia*) *trichopyga*	F	French Guiana	Saint-Georges	02/10/07	KU761711
131B	*L.* (*Viannamyia*) *furcata*	F	French Guiana	Cacao	10/12/07	KU761638
138B	*L.* (*Viannamyia*) *furcata*	F	French Guiana	Cacao	10/12/07	KU761641
29B	*L.* (*Viannamyia*) *furcata*	F	French Guiana	Cacao	10/12/07	KU761817
1206A	*L.* (*Viannamyia*) *tuberculata*	F	French Guiana	Cacao	19/04/07	KU761632
2649B	*L.* (*Viannamyia*) *tuberculata*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2674B	*L.* (*Viannamyia*) *tuberculata*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2680B	*L.* (*Viannamyia*) *tuberculata*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2684B	*L.* (*Viannamyia*) *tuberculata*	F	French Guiana	Saint-Georges	02/10/07	KU761711
2905B	*L.* (*Viannamyia*) *tuberculata*	F	French Guiana	Saint-Georges	02/10/07	KU761802
2908B	*L.* (*Viannamyia*) *tuberculata*	F	French Guiana	Saint-Georges	02/10/07	KU761803
2909B	*L.* (*Viannamyia*) *tuberculata*	F	French Guiana	Saint-Georges	02/10/07	KU761804

Appendix 3: DNA extraction protocols for pools of sand flies

DNA extraction protocols for pools of sand flies

Prior to DNA extraction individuals were pooled in 2 mL microcentrifuge tubes with an autoclaved glass bead. The samples were then stored at -80°C for 30 min and crushed using Qiagen TissueLyser 2 for 30 s.

Qiagen extraction kit

We used the DNeasy Blood and Tissue kit following the manufacturer protocol with modified quantities (500 µL Buffer ATL, 50 µL of 20 mg/mL Proteinase K, 500 µL Buffer AL, 500 µL absolute ethanol, 500 µL Buffer AW1, 500 µL Buffer AW2, elution in 200 µL Buffer AE).

Chelex protocol

We used the Chelex without boiling protocol described by (Casquet *et al.* 2012) with 450 µL of 10% Chelex solution and 30 µL of 20 mg/mL Proteinase K in each sample.

Appendix 4: DNA reference database in OBITools extended fasta format and modified NCBI taxonomic database in ecopcr format

Available online

Appendix 5: Metabarcoding data for mock communities and field samples.

We included the sequencing data in fasta format after dereplication and removal of singletons, as well as in tabular format after taxonomic assignment of the sequences. The sample description files used for the assignment of the sequences to their corresponding samples with the OBITools are also included.

Available online

Appendix 6: Bash script used for the bioinformatic treatment of the sequencing data

```
################################################################################
##Bash script used for the bioinformatic treatment of the sequencing     ##
##data with the OBITools package (see the complete OBITools documentation##
##at (http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12428/pdf).   ##
##The first part concerns the generation of the DNA reference database,  ##
##and the second  concerns the treatment of the metabarcoding data.      ##
################################################################################

#### DNA REFERENCE DATABASE ####

# Alignment and merging of paired-end reads using illuminapairedend.
# The program assigns an alignment score to each resulting sequence based
# on the phred quality scores and the length of the aligned regions.

illuminapairedend --fasta-output -r readsR1.fastq readsR2.fastq \
> readsR1R2.fasta

# Reads assignment using ngsfilter. The program requires a table providing
# the information regarding the primer pair and the tag combination used
# for each sample (see the OBITools documentation for more details). This
# step will add an attribute to each sequence containing the name of the
# corresponding sample. Other information such as the scientific name of
# the taxa can be added.

ngsfilter -t ngsfilter.tab -e 2 --nuc readsR1R2.fasta > \
readsR1R2_ngsfilt.fasta

# Removal of low quality reads using obigrep (alignment scores<50,
# containing Ns or shorter than 50bp)

obigrep -s '^[acgt]+$' -l 50 -p 'score>=50' readsR1R2_ngsfilt.fasta \
> readsR1R2_ngsfilt_lowqual.fasta

# Dereplication of the sequences using obiuniq (regroups every identical
# reads assigned to the same sample into one sequence and keeps the
# coverage information)

obiuniq -c sample readsR1R2_ngsfilt_lowqual.fasta \
> readsR1R2_ngsfilt_lowqual_derep.fasta

# Removal of the sequences supported by less than 50 reads using obigrep

obigrep -p 'count>50' readsR1R2_ngsfilt_lowqual_derep.fasta \
> readsR1R2_ngsfilt_lowqual_derep_min50.fasta

# Selecting the majority sequence for each sample using obiselect

obiselect -n 1 -c sample -f count -M \
readsR1R2_ngsfilt_lowqual_derep_min50.fasta \
> readsR1R2_ngsfilt_lowqual_derep_min50_maj.fasta
```

```
# convert NCBI taxonomic database (release 197) in ecopcr format using
# obiconvert

obiconvert --genbank -t taxdump/ --ecopcrDB-output=ncbi_r197

# Add all taxa included in the reference database into the ncbi taxonomic
# database using obitaxonomy. The reference database fasta file file must
# be first annotated with an attribute containing the name of the taxon and
# an attribute containing the taxonomic path of the sample. In our case, we
# further modified the taxonomic database using obitaxonomy to make it fit
# the classification of Young & Duncan (1994) (command lines not presented
# here, see the Obitools manual for further detail concerning the use of
# obitaxonomy).

obitaxonomy -d ncbi_r197 -p path -k taxonomy -F \
readsR1R2_ngsfilt_lowqual_derep_min50_maj.fasta

# Add taxids in the attributes of the sequences using obiaddtaxids. The
# program requires an attribute in the sequence containing the scientific
# name of the sample as well as an ecopcr formatted taxonomic database

obiaddtaxids -k taxonomy -d ncbi \
readsR1R2_ngsfilt_lowqual_derep_min50_maj.fasta > phleb_DB_FG.fasta

#The resulting reference database is ready to be used for taxonomic
# assignations of samples with ecotag

#### METABARCODING DATA ####

# All steps until sequence dereplication are the same as previously
# described

# Removal of singletons using obigrep

obigrep -p 'count>1' readsR1R2_ngsfilt_lowqual_derep.fasta \
> readsR1R2_ngsfilt_lowqual_derep_noS.fasta

# Initial denoising:
# (i) dividing the fasta file into separate files for each sample using
# obisplit.
# (ii) clustering of the sequences within each sample using sumaclust with
# a similarity threshold of 0.99
# (iii) selection of the majority sequence in each cluster using obiselect
# (iv) merging the resulting files

obisplit -p obisplit_ -t readsR1R2_ngsfilt_lowqual_derep_noS.fasta

for i in obisplit_* ;do sumaclust -t 0.99 $i \
> $(echo $i | sed 's/.fasta/_cl_99.fasta/');done

for i in *_cl_99.fasta; do obiselect -n 1 -c cluster -f count -M $i \
> $(echo $i | sed 's/_cl_99.fasta/_denoised.fasta/'); done

cat *_denoised.fasta > readsR1R2_ngsfilt_lowqual_derep_noS_denoised.fasta
```

```
rm obisplit_*

# Taxonomic assignment of the sequences using ecotag. The program requires
# a reference database as well as the taxonomic files of an ecoPCR database
# (.ndx, .rdx and .tdx files) obtained as explained previously.

ecotag -R phleb_DB_FG.fasta -d ncbi_r197 \
readsR1R2_ngsfilt_lowqual_derep_noS_denoised.fasta \
> readsR1R2_ngsfilt_lowqual_derep_noS_denoised_ecotag.fasta

# output in tabular format using obitab

obitab -d -o readsR1R2_ngsfilt_lowqual_derep_noS_denoised_ecotag.fasta \
> readsR1R2_ngsfilt_lowqual_derep_noS_denoised_ecotag.tab
```

Appendix 7: R script used for the filtering of metabarcoding data and output as community matrices

```
#################################################################
##Sandfly metabarcoding: sequence filtering and results output##
#################################################################

# This script is to be used on metabarcoding data in tabular format
# generated with the obitools (http://metabarcoding.org/obitools/doc/). The
# script uses sequenceattributes that are added during the obitools
# pipeline. They are stored by columns, and correspond to the following:
# - sample: name of the sample to which the sequence has been assigned
# -count: sequence reads count in each sample
# -best_identity.phleb_DB_FG: percentage of identity with the best match in
# the reference database "phleb_DB_FG"
# -scientific_name: name of the taxa to which the sequence has been
# assigned
# -sequence: the sequence itself
# The script also uses custom attributes in the ngsfilter tab:
# -DNA_sample: name of the DNA sample from wich the sequence was amplified
# (this differ from the "sample" attribute by the fact that several PCR
# replicates were done on each DNA sample)
# -PCR_replicate_id: PCR replicate identifier
# The mistags-based filtering relies on assignation of sequences to
# non-used tag combosfor wich the DNA_sample attribute has been set to "NU"
# in the ngsfilter tab.

##### DATA IMPORT (obitab output) ####

data=read.table("mock_communities_R1R2_ngsfilt_lowqual_derep_noS_denoised_e
cotag.tab",sep='\t',header=T,stringsAsFactors = F)

##### SEQUENCE FILTERING ####

# reference-based filtering (RBF)
# ===============================
# Keeps only the sequences that have at least 97% identity whith its best
# match in  the reference database

data_RBF = data[data$best_identity.phleb_DB_FG>0.97,]

# mistags-based filtering (MBF)
#==============================
# We use a procedure similar to what proposed by Esling et al. (2015) to
# remove sequences resulting from tag switching events, which cause false
# positives. The abundance of sequences assigned to non-used tag combos
# (NUs, one tag combination over two in a latin square design) is used as
# an indicator of the frequency of mistags due to tag-switching events. For
# each OTU, a threshold is defined as twice the 95% quantile of the
# abundance of the OTU in all NUs. Then, for each sample, the OTU is kept
# only if its abundance is greater than its corresponding treshold.

# computation of the mistags-based thesholds per OTU
```

```
OTU_threshold = 2*tapply(data_RBF$count[data_RBF$DNA_sample=="NU"],
data_RBF$scientific_name[data_RBF$DNA_sample=="NU"],
function(x)quantile(x,0.95,na.rm=T))

# for each sample, remove OTUs having an abundance lower than its
# corresponding threshold

data_RBF_MBF = NULL

for (i in 1:nrow(data_RBF))
{
  if (data_RBF$count[i] > OTU_threshold[data_RBF$scientific_name[i]] |
      !data_RBF$scientific_name[i] %in% names(OTU_threshold))
    data_RBF_MBF = rbind(data_RBF_MBF,data_RBF[i,])
}

# PCR replicate filtering (PCRrep):
# ================================
# remove sequences that are not found in at least 2 PCR replicates for each
# DNA extraction

data_RBF_MBF_PCRrep = NULL

for (i in 1:nrow(data_RBF_MBF))
{
  if (length(unique(data_RBF_MBF$PCR_replicate_id[data_RBF_MBF$sequence ==
data_RBF_MBF$sequence[i] &

data_RBF_MBF$DNA_sample==data_RBF_MBF$DNA_sample[i]])) >= 2)
    data_RBF_MBF_PCRrep = rbind(data_RBF_MBF_PCRrep,data_RBF_MBF[i,])
}

##### RESULTS OUTPUT ####

# merge the data: sum the sequence reads counts of each reference-based
# OTUs in each DNAsample and format as a community dataset (species in
# columns, samples in rows).

merged_counts=aggregate(data_RBF_MBF_PCRrep$count,data_RBF_MBF_PCRrep[,c("D
NA_sample","scientific_name")],FUN=sum)

community_readCounts=as.data.frame.matrix(xtabs(merged_counts$x~merged_coun
ts$DNA_sample+merged_counts$scientific_name))

# create a presence/absence matrix (PA)

community_PA = ifelse(community_readCounts>0, 1, 0)

# create a matrix with the relative reads abundances in each sample (RA)

community_RA = community_readCounts/rowSums(community_readCounts)

# Note:
# Presence/absence data can be summed accross samples to caracterize
# communities at largerscales. When normalized with the number of samples
```

```
# in a given area, this would return toits occurence frequency in that
# area.This appraoch is also applicable to weighted relative abundance
# data, assuming that PCR amplificaiton bias are consistentaccross samples.
# In contrary, it would be inappropriate to sum raw reads counts accross
# samples, as they are not necessarily indicative of the template DNA
# quantities prior PCR amplification.
```

PRELIMINARY RESULTS:
MODIFICATION OF SAND FLY COMMUNITIES IN VERTEBRATE-DEPLETED HABITATS.

INTRODUCTION

Characterizing the impacts of species loss on ecosystem services is an urgent need in an era marked by biodiversity erosion (Cardinale *et al.* 2012). In particular, the impoverishment of vertebrate fauna may have various consequences on ecosystem dynamics and diversity (Dirzo *et al.* 2014) due to cascading effects on other organisms through the alteration of trophic interactions (Myers *et al.* 2007; Kurten 2013; Keesing & Young 2014). Hematophagous arthropods rely on the availability of vertebrate hosts in their environment as blood meal sources, and may therefore be affected by defaunation. Since many of them are vectors of infectious diseases, the impact of vertebrate fauna modifications on hematophagous communities may have indirect epidemiological implications. Specifically, the link existing between host diversity, vector abundance and circulation of parasites is a central question in recent biodiversity-disease research (Randolph & Dobson 2012; Ostfeld & Keesing 2012; Johnson *et al.* 2015).

The extent to which hematophagous arthropod rely on local vertebrate populations is driven by ecological features of the taxa. Blood-sucking ticks (Ixodoidea), for instance, are very dependent on host abundance and movement for feeding opportunities due to their "site-and-wait" behaviour (Ogden & Tsao 2009; Randolph & Dobson 2012). On the contrary, highly mobile mosquitoes (Culicidae) are thought to be limited by the availability of breeding sites rather than that of blood meal sources, which is a classical assumption in mosquito-borne disease modelling (Dobson 2004; Ogden & Tsao 2009). In addition, hematophagous arthropod's feeding preferences will likely determine their resilience to vertebrate community modifications: opportunistic feeders have less chance to be affected by a decline in host diversity than more specialized ones (Clavel *et al.* 2011).

Sand flies (Diptera, Psychodidae, Phlebotominae) are small dipterans responsible for the transmission of leishmaniases as well as other viral and bacterial diseases worldwide (Depaquit *et al.* 2010; Ready 2013). Individuals of both sex acquire sugar compounds by feeding on plants and insect honeydew (Schlein & Muller 1995; Muller & Schlein 2004). Additionally, females require blood meals to complete the development of their egg batches, although some species are known to display autogeny (*i.e.* the ability to complete egg development without blood meal, *e.g.* Montoya-Lerma 1992; Dinesh *et al.* 2008). Some sand fly species are relatively generalist, while others exhibit marked feeding preferences (Tesh *et al.* 1971; Christensen *et al.* 1982;). Furthermore, in comparison to mosquitoes, sand flies have

modest flight ranges that may not allow them to actively seek for vertebrate hosts on large areas (Morrison *et al.* 1993; Casanova *et al.* 2005; Orshan *et al.* 2016). Therefore, it can be expected that both vertebrate diversity and abundance will affect sand fly communities. However, this question has not been directly addressed. Most studies aiming at explaining sand fly species distribution have rather focused on the effect of large-scale environmental variables (Peterson & Shaw 2003; Moo-Llanes *et al.* 2013), or on that of plant and abiotic habitat structure (e.g. Rotureau *et al.* 2006; Claborn *et al.* 2008; Nieves *et al.* 2014).

The aim of this work was to explore the effect of human-induced vertebrate fauna depletion on sand fly communities. The study was conducted in the Amazonian rainforest of French Guiana. were zoonotic leishmaniasis is endemic. Sand flies were collected in sites undergoing contrasting levels defaunation and analysed using metabarcoding to characterize community composition.

MATERIAL AND METHODS

Sampling

Sand flies were collected in two localities in French Guiana: in the area of Counami near the city of Iracoubo in October 2015 and in the area of Saint-Georges de l'Oyapock in March 2016. In each locality, two forest sites of *c.* one hectare were selected. Within localities, sites were separated by less than 25 km, harboured similar forest habitat (as defined by Guitet *et al.* 2015), but were subject to contrasting levels of anthropogenic pressure. Anthropogenic pressure was defined by a previously published human footprint index (de Thoisy *et al.* 2010) and the distance to the closer inhabited area (Table 1). The human footprint index has been shown to correlate with lower vertebrate diversity in French Guiana (de Thoisy *et al.* 2010, 2016). In particular, in Counami, line transect censuses of *c.* 100 cumulated km were conducted near sampling sites in the frame of these previous studies and indicated lower diversity and abundance of large vertebrates in the most perturbed site. In Saint-Georges, depletion of the vertebrate fauna on the most perturbed site was further suggested by blood meal analyses of sand flies and mosquitoes (see next chapter), as well as dung beetle diversity and camera-trap results (F. Feer and P.-M. Forget, personal communication).

In each site, center of disease control light traps were set during three to four consecutive nights (from 6 pm to 6am). After each night, insects contained in the traps were killed by freezing. Sand fly females were then directly sorted and pooled by trap-night in

microcentrifuge tubes with 95% ethanol until laboratory work. A maximum of 50 individuals was included in a pool. When a trap contained more than 50 individuals, several pools were made.

Table 1: Sampling sites along with the criteria used for the definition of anthropogenic pressure levels

Locality	Site	Long.	Lat.	HFP*	DIA (km)**	Relative perturbation
Counami	CO-	-53.27215	5.38337	13	28	-
	CO+	-53.17763	5.40939	17	4	+
Saint-Georges	SG-	-51.98967	4.03493	21	28	-
	SG+	-51.81871	3.88359	32	2	+

*Human footprint index: based on that of de Thoisy (2010) and further updated in 2012
**Distance to the closest inhabited area following roads or tracks

DNA amplification and sequencing

From each trap-night, a maximum of 200 specimens (*i.e.* four pools) were used for laboratory work. We extracted DNA from each engorged female using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA). The Ins16S_1 marker was amplified (Ins16S_1-F: 5'-TRRGACGAGAAGACCCTATA-3', Ins16S_1-R: 5'-TCTTAATCCAACATCGAGGTC-3'; Clarke *et al.* 2014), with three PCR replicates, in order to identify sand fly species contained in the pools by metabarcoding (Kocher *et al.* 2017). Blood-fed specimens were barcoded individually using the same marker. Tags of eight base pairs with at least five differences between them were added at the 5' end of each primer to enable the sequencing of the multiple PCR products in a single sequencing run (Binladen *et al.* 2007). Only one out of two possible tag combinations was used (*Latin square design*; see Esling *et al.* 2015), in order to perform mistag-based filtering of the sequences.

PCR products were pooled and sent for library construction and sequencing to the GeT-PlaGe core facilities of Genotoul (Toulouse, France). Samples were diluted in ultrapure water. A volume of 130 µL containing 3 µg of DNA was purified using the HighPrep PCR system (Magbio Genomics, Gaithersburg, MD, USA) and used for library construction with the Illumina NEXTflex PCR-Free DNA sequencing kit following the instructions of the supplier (Bioo Scientific corp., Austin, TX, USA). Purified fragments were end-repaired, A-tailed and ligated to sequencing indexed adapters. The quality of the library was controlled using the Fragment Analyzer (Advanced Analytical, Ames, IA, USA) and quantified by qPCR with the Library Quantification Kit - Illumina Genome Analyzer-SYBR Fast Universal (CliniSciences, Nanterre, France). The library was pooled with that of other projects and loaded onto the Illumina MiSeq cartridge according to the manufacturer instructions. The quality of the run

was checked internally using PhiX. Quality filtering was performed by the Consensus Assessment of Sequence and Variation (CASAVA) pipeline. The sequencing data was stored on the NG6 platform (Mariette *et al.* 2012) and all computations were performed on the computer cluster of the Genotoul bioinformatic platform (Toulouse, France).

Bioinformatic treatment and taxonomic assignations

The sequencing data was analysed using the `OBITOOLS` package (Boyer *et al.* 2016), as described previously (Kocher *et al.* 2017). Brefly, pair-end reads were aligned, merged and then assigned to their corresponding sample based on the tagged primer sequences with two mismatches allowed. Low quality reads were removed and reads were then dereplicated. Sequences represented by only one read (singletons) were discarded. Taxonomic assignations were performed using `ecotag`. We then applied reference based (97% threshold), mistag-based, and PCR replicate-based filtering as described in (Kocher *et al.* 2016). Eventually, species abundances in samples were estimated by weighting the relative abundance of reads assigned to each species by the number of individuals contained in the sample.

Sand fly community analyses

Some trap-nights contained very low or high number of sand flies compared to the average catch on the same site, which we suspected to be due to dysfunction of the material or unusual biological events (such as massive emergence of sand fly adults next to the trap). To define outlier traps, we log-transformed sand fly abundances and computed modified Z-scores within each site, as suggested by (Iglewicz & Hoaglin 1993). If the modified Z-score was greater than 3.5, the trap-night was removed from data prior to remaining analyses.

Mean sand fly abundance per trap-night and overall proportion of blood-fed specimens was computed for each site. Confidence intervals were calculate using the Poisson exact method implemented in the epitools R package for mean abundances (Aragon *et al.* 2010) and the classical normal approximation for proportions of blood-fed specimens.

Sand fly species distribution in trap-nights were compared using Hellinger distance (Legendre & De Cáceres 2013). The distance matrix was reduced in a 2D-space by using non-metric multidimensional scaling (NMDS), as implemented in the package R 'vegan' (Dixon 2003; Legendre & Legendre 2012). Analysis of similarity (ANOSIM) was performed to statistically assess of the similarity of trap-night contents within localities or level of perturbation.

Trap-night contents were summed in each site to obtain a community matrix. In order to compare sand fly diversity between sites, we estimated Hill numbers of diversity and corresponding 95% confidence intervals along rarefied and extrapolated curves as suggested by Chao *et al.* (2014), using the R package iNEXT (Hsieh *et al.* 2016).

RESULTS

A total of 8,700 female sand flies, including 98 blood fed specimens (1.12%), were collected in 129 trap-nights. After computing modified Z-scores, one trap containing only one specimen in SG- and another containing 1896 specimens in SG+ were considered as outliers and removed from the dataset. Thus, the remaining sample contained 6.803 specimens.

The average number of specimens caught per trap-night was 86.1 in Saint-Georges and 16.1 in Counami. In both localities, mean sand fly abundance was significantly higher in the most perturbed site (42.2/135.4 in SG-/SG+, 12.5/19.7 in CO-/CO+; Figure 1A). On the contrary, the proportion of blood fed specimens was lower in the most perturbed sites (1.6%/1.1% in SG-/SG+, 3.5%/2.3% in CO-/CO+; Figure 1B). This difference was not significant when comparing both pairs of site separately, but was significant on the overall (X^2=4.6, p-value=0.03).

In total, metabarcoding allowed the detection of 36 species (or MOTUs) of sand flies. Estimations based on relative abundances of sequencing reads indicated that the most abundant species were *Trichopygomyia trichopyga* (46.5%), *Trichophoromyia ininii* (12.9%) and *Psychodopygus s. maripaensis* (10.2%). Among blood fed specimens, the most abundant species were *Nyssomyia umbratilis* (27.9%), *Psychodopygus ayrozai* (18.6%), *Ps. maripaensis* (9.3%). Only two blood-fed *Ty. trichopyga* and no blood-fed *Th. ininii* were found which was clearly unexpected regarding the abundance of these species and the overall proportion of blood fed specimens (Fisher's exact test p-value = 2.2×10^{-16} and 1.1×10^{-5} respectively).

NMDS ordination highlighted differences in trap contents between locality (Figure 2A), which was further attested by ANOSIM result (R=0.24, p<0.001). Trap contents collected in the most perturbed sites also seemed to be grouped in NMDS ordination. Indeed, ANOSIM showed that the mean ranked dissimilarity was lower within most perturbed sites than within less perturbed sites or between these two groups (Figure 2B).

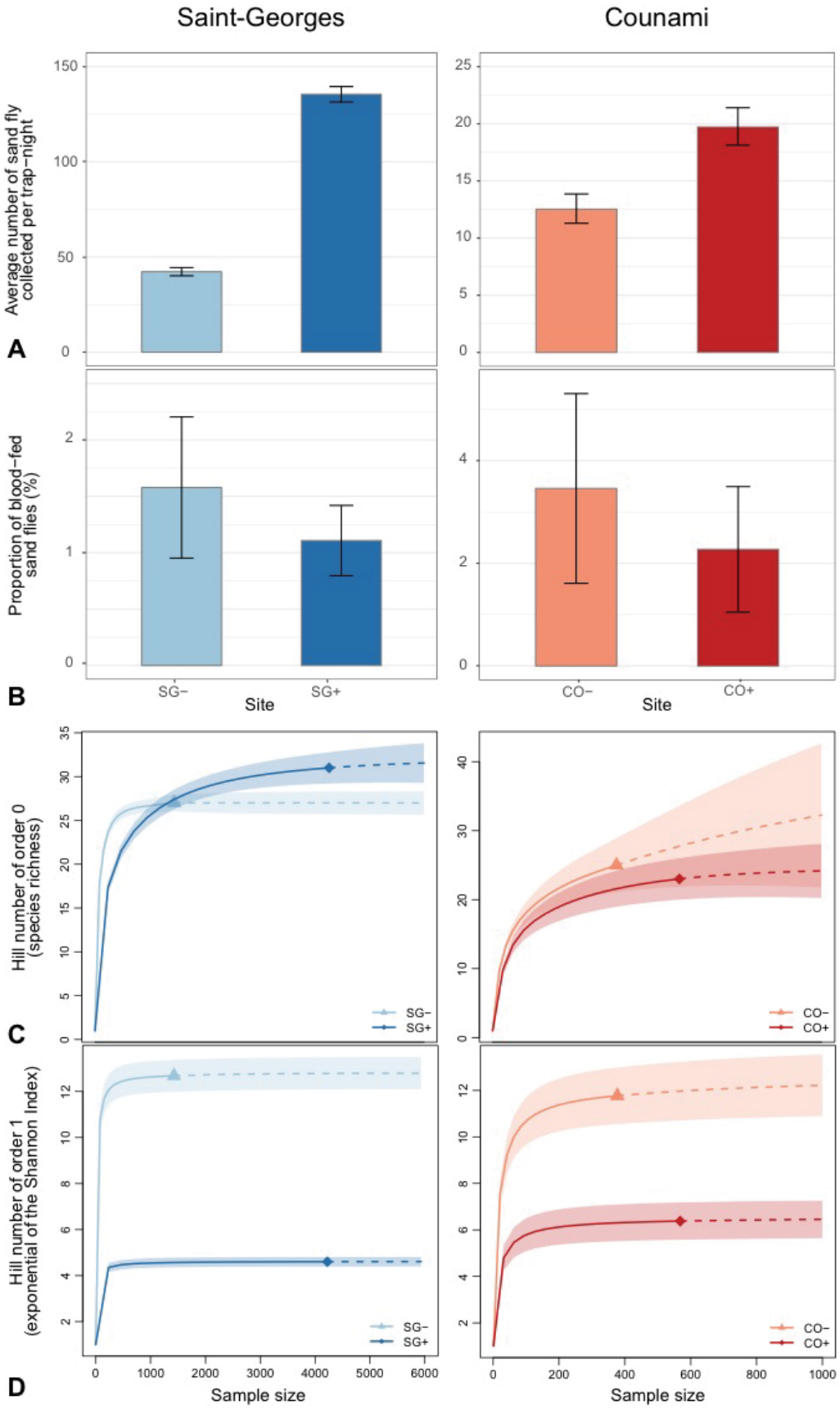


Figure 1: Comparison of mean sand fly abundance **(A)**, proportion of blood-fed sand flies **(B)**, and sand fly diversity rarefaction (continuous lines)/extrapolation (dotted lines) curves **(C)** and **(D)** found in less versus most perturbed site within each sampling locality.

Species accumulation curves indicated that a large proportion of species were sampled in each site (Figure 1C). Twenty-seven species were detected in SG-, 31 in SG1+, 25 in CO- and 23 in CO+. In each locality, species richness showed poorly significant differences along the rarefaction/extrapolation curves between sites. On the contrary, Hill numbers of order 1 (*i.e.* exponential of the Shannon Index) indicated a marked decrease in the most impacted sites, which revealed lower homogeneity in species relative abundances. Indeed, *Ty. trichopyga* and *Th. ininii* were found to be highly dominant in SG+, and *Ty. trichopyga* was also found as a largely dominant species in CO+ (Figure 3).

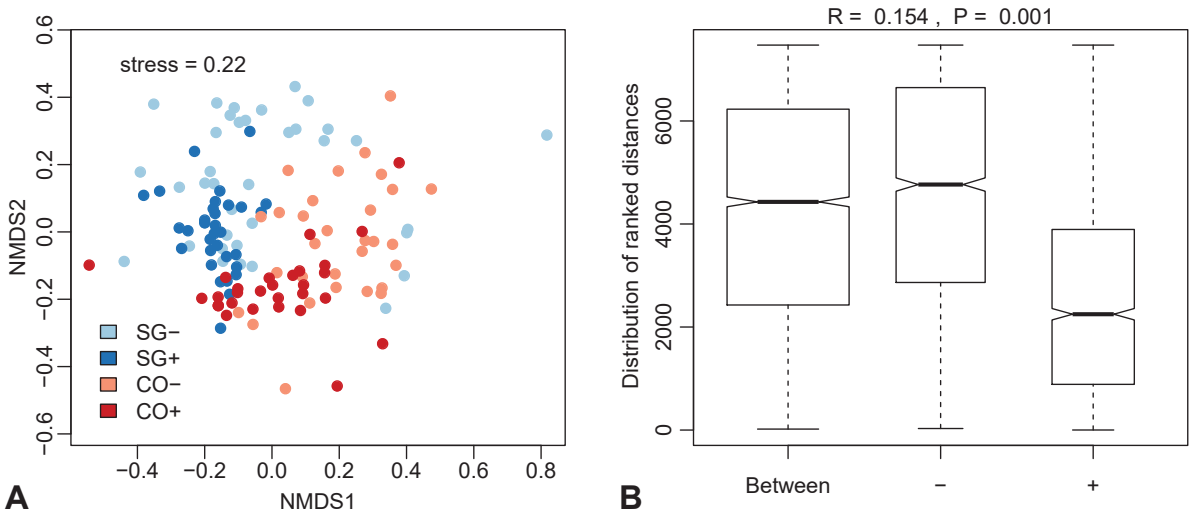


Figure 2: (A) NMDS ordination of trap sand fly species contents. **(B)** Distribution of ranked dissimilarities found within less perturbed sites (-), most perturbed sites (+) and between these two categories.

DISCUSSION

There is a growing concern about the consequences of ecosystem alteration on human health (Myers *et al.* 2013). The “dilution effect” hypothesis states that more diverse communities are less prone to pathogen transmission due to the presence of poorly competent hosts (Ostfeld & Keesing 2012). However, the generality of such phenomenon is still under debate (Wood *et al.* 2014; Levi *et al.* 2016). For vector borne-disease, an opposite "amplification" effect can be observed if host diversity also promotes abundance of competent vectors (Randolph & Dobson 2012).

In this work, we explored the effect of human-induced vertebrate fauna depletion on sand fly communities, with a metabarcoding approach. The comparison of the results obtained in two distant localities revealed congruent patterns (Figure 1):

- sand fly abundance was higher in the most perturbed sites

- Species richness did not vary much between sites, whereas species abundances homogeneity (as measured by the Hill number of order 1) decreased dramatically in the most perturbed site.

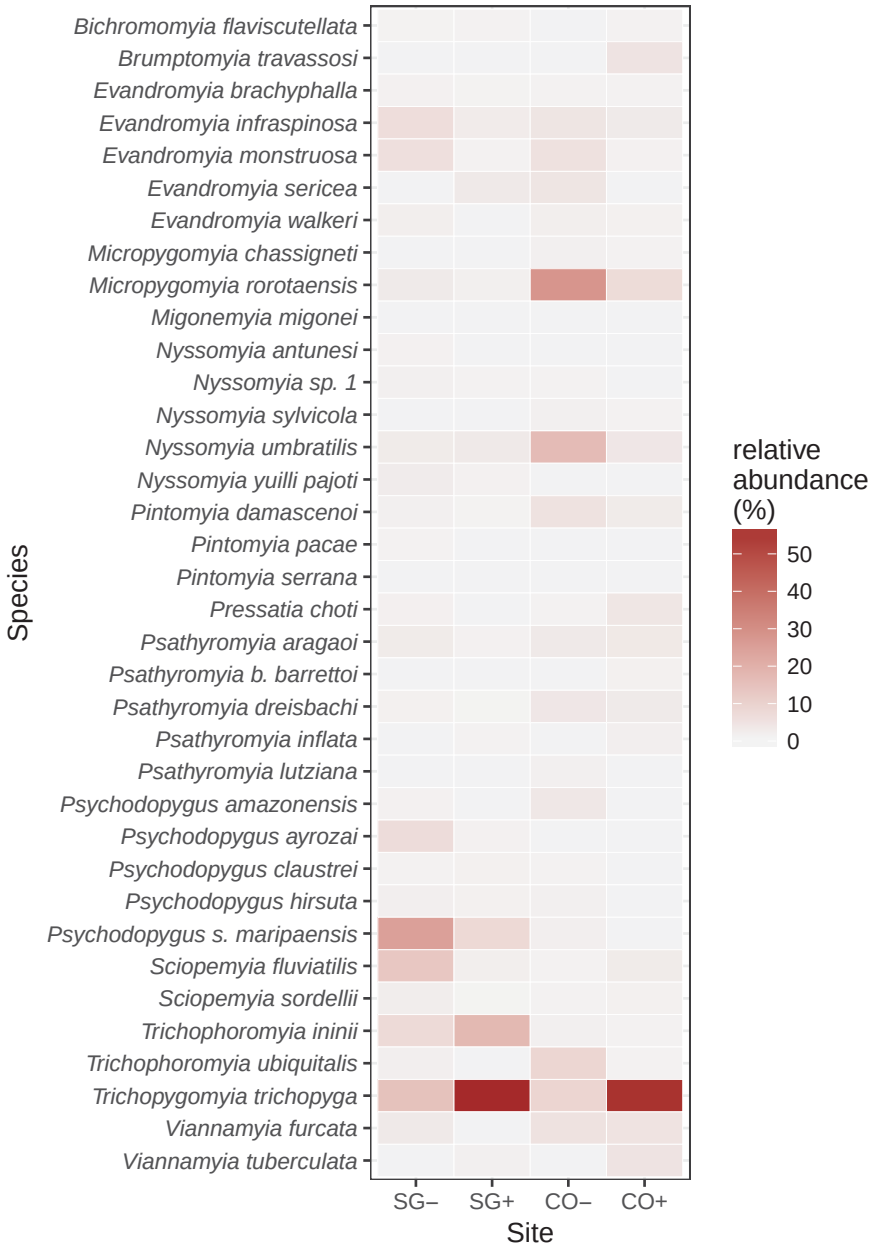


Figure 3: Heatmap representing sand fly community composition found in the four sampling sites

These patterns resulted from the very high abundance of one or two hyper-dominant species in the most perturbed sites (*Ty. trichopyga* in CO+, *Ty. trichopyga* and *Th. ininii* in SG+). *Ty. trichopyga* is not known as a vector of leishmaniasis and *Th. ininii* was only once reported infected by an unidentified *Leishmania* in the region (Fouque et al. 2007). Therefore, the epidemiological consequences of these observations are not clear. A striking observation was that almost no blood fed specimens were found among individuals of these two species,

despite that they represented the majority of collected sand flies. This resulted in a significant decrease of the overall proportion of blood-fed specimens in perturbed sites. Some sand fly species are known be able to develop their ovaries without a blood meal (autogeny; Montoya-Lerma 1992; Dinesh *et al.* 2008). If *Ty. trichopyga* and *Th. ininii* were particularly prone to autogeny, this could explain the remarkably low proportion of blood-fed specimens found in these species as well as their better resilience in vertebrate-depleted sites. Why an autogenic species would thrive, rather than just resist, in an environment where blood meal sources are scarce, seems less evident. One could think of indirect processes, such as a relief from competition with other less autogenic species. However, our data does not really support this hypothesis since the observed increase in *Ty. trichopyga* and *Th. ininii* abundance in the most perturbed sites were not compensated by a rarefaction of other (supposedly less autogenic) species, resulting in an important increase of total sand fly abundance.

Blood meal source depletion is not the only possible consequence of defaunation that could affect sand fly communities. In particular, modifications of vertebrate fauna may impact plant habitats through different interactions such as seed dispersal and predation as well as herbivory or trampling (Kurten 2013; Dirzo *et al.* 2014). Indeed, the vegetation represents another feeding resource for sand flies, in addition to providing them reproduction and resting sites (Schlein & Muller 1995; Feliciangeli 2004; Lima *et al.* 2016).Additionally, vertebrate burrows are known to house various sand fly species (Chippaux & Pajot 1983; Feliciangeli 2004; Rotureau *et al.* 2006). Furthermore, defaunation may be difficult to disentangle from other type of human-induce disturbances because these are likely to happen in the same areas. For instance, even though we aimed at minimizing environmental variability between our sampling sites, logging was known to have occurred in CO+. Lower sand fly community diversity and higher dominance in more perturbed habitats was previously observed (Nieves *et al.* 2014), and was also found here. However, the addition of more study sites and more detailed characterization of habitats would be required to confirm its generality and state on its precise cause.

Sand fly populations may fluctuate in time with peaks depending on the species and local climatic conditions (Shaw & Lainson 1972; Le Pont 1982; Oliveira *et al.* 2008). It cannot be ruled out that synchronized adult emergences of these species happened by chance on the most perturbed sites during sampling periods. This may explain that almost no blood-fed specimens were found for these species since freshly borne females should have less chance to have fed on a vertebrate host when caught. This would also explain that the very large numbers of *Ty. trichopyga* and *Th. ininii* observed in the most perturbed sites were not

compensated by a decrease in other species abundance, but seemed rather added to a "base" community. Nevertheless, the same pattern was observed in two very distant localities sampled during distinct period of the year. This do suggest the existence of another underlying mechanism. However, whether these observations are directly related to defaunation and have epidemiological consequences are questions that necessitate further work.

REFERENCES

Aragon, T.J., Enanoria, W.T., Fay, M., Porco, T. & Samuel, M. (2010). *Applied epidemiology using R*. MedEpi Publishing. http://www. medepi. net/epir/index. html. Calendar Time. Accessed.

Binladen, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R. & Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, **2**, e197.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. & Coissac, E. (2016). obitools: a unix-inspired software package for DNA metabarcoding. *Molecular ecology resources*, **16**, 176–182.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S. & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, **486**, 59–67.

Casanova, C., Costa, A.I. & Natal, D. (2005). Dispersal pattern of the sand fly *Lutzomyia neivai* (Diptera: Psychodidae) in a cutaneous leishmaniasis endemic rural area in Southeastern Brazil. *Memórias do Instituto Oswaldo Cruz*, **100**, 719–724.

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K. & Ellison, A.M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45–67.

Chippaux, J.-P. & Pajot, F.-X. (1983). La leishmaniose en Guyane française. IV: Note préliminaire sur les Phlébotomes des terriers. *Cahiers - ORSTOM. Entomologie médicale et parasitologie*, **21**, 149–154.

Christensen, H.A., Arias, JR, A., de Vasquez, AM, de V. & de Freitas, RA, de F. (1982). Hosts of sandfly vectors of *Leishmania braziliensis guyanensis* in the central Amazon of Brazil. *The American journal of tropical medicine and hygiene*, **31**, 239–242.

Claborn, D., Masuoka, P., Morrow, M. & Keep, L. (2008). Habitat analysis of North American sand flies near veterans returning from *Leishmania*-endemic war zones. *International Journal of Health Geographics*, **7**, 65.

Clarke, L.J., Soubrier, J., Weyrich, L.S. & Cooper, A. (2014). Environmental metabarcodes for insects: *in silico* PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, **14**, 1160–1170.

Clavel, J., Julliard, R. & Devictor, V. (2011). Worldwide decline of specialist species: toward a global functional homogenization? *Frontiers in Ecology and the Environment*, **9**, 222–228.

Depaquit, J., Grandadam, M., Fouque, F., Andry, P.E. & Peyrefitte, C. (2010). Arthropod-borne viruses transmitted by Phlebotomine sandflies in Europe: a review. *Euro Surveill*, **15**, 19507.

Dinesh, D.S., Kumar, V., Kesari, S., Kumar, A.J., Das, P. & others. (2008). Is *Phlebotomus argentipes* Annandale and Brunetti (Diptera: Psychodidae) autogenous? *Journal of vector borne diseases*, **45**, 174.

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J. & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, **345**, 401–406.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Dobson, A. (2004). Population Dynamics of Pathogens with Multiple Host Species. *The American Naturalist*, **164**, S64–S78.

Esling, P., Lejzerowicz, F. & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, **43**, 2513–2524.

Feliciangeli, M.D. (2004). Natural breeding places of phlebotomine sandflies. *Medical and Veterinary Entomology*, **18**, 71–80.

Fouque, F., Gaborit, P., Issaly, J., Carinci, R., Gantier, J.-C., Ravel, C. & Dedet, J.-P. (2007). Phlebotomine sand flies (Diptera: Psychodidae) associated with changing patterns in the transmission of the human cutaneous leishmaniasis in French Guiana. *Memórias do Instituto Oswaldo Cruz*, **102**, 35–40.

Guitet, S., Brunaux, O., de Granville, J. & Richard-Hansen, C. (2015). *Catalogue des habitats forestiers de Guyane*. ONF and DEAL, Cayenne.

Hsieh, T.C., Ma, K.H. & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers) (G. McInerny, Ed.). *Methods in Ecology and Evolution*.

Iglewicz, B. & Hoaglin, D.C. (1993). *How to detect and handle outliers*. Asq Press.

Johnson, P.T.J., Ostfeld, R.S. & Keesing, F. (2015). Frontiers in research on biodiversity and disease. *Ecology Letters*, **18**, 1119–1133.

Keesing, F. & Young, T.P. (2014). Cascading Consequences of the Loss of Large Mammals in an African Savanna. *BioScience*, **64**, 487–495.

Kocher, A., Gantier, J.-C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., Dusfour, I., Girod, R., Bañuls, A.-L. & Murienne, J. (2017). Vector soup: high-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources*, **17**, 172–182.

Kurten, E.L. (2013). Cascading effects of contemporaneous defaunation on tropical forest communities. *Biological Conservation*, **163**, 22–32.

Le Pont, F. (1982). La leishmaniose en Guyane française: 2. Fluctuations saisonnières d'abondance et du taux d'infection naturelle de *Lutzomyia* (*Nyssomyia*) *umbratilis* Ward et Fraiha, 1977. *Cahiers ORSTOM. Série Entomologie Médicale et Parasitologie*, **20**, 269–277.

Legendre, P. & De Cáceres, M. (2013). Beta diversity as the variance of community data: dissimilarity coefficients and partitioning (H. Morlon, Ed.). *Ecology Letters*, **16**, 951–963.

Legendre, P. & Legendre, L.F. (2012). *Numerical ecology*. Elsevier.

Levi, T., Massey, A.L., Holt, R.D., Keesing, F., Ostfeld, R.S. & Peres, C.A. (2016). Does biodiversity protect humans against infectious disease? Comment. *Ecology*, **97**, 536–542.

Lima, L.H.G. de M., Mesquita, M.R., Skrip, L., de Souza Freitas, M.T., Silva, V.C., Kirstein, O.D., Abassi, I., Warburg, A., Balbino, V. de Q. & Costa, C.H.N. (2016). DNA barcode for the identification of the sand fly *Lutzomyia longipalpis* plant feeding preferences in a tropical urban environment. *Scientific Reports*, **6**, 29742.

Mariette, J., Escudié, F., Allias, N., Salin, G., Noirot, C., Thomas, S. & Klopp, C. (2012). NG6: Integrated next generation sequencing storage and processing environment. *BMC genomics*, **13**, 462.

Montoya-Lerma, J. (1992). Autogeny in the Neotropical Sand Fly *Lutzomyia lichyi* (Diptera: Psychodidae) from Colombia. *Journal of Medical Entomology*, **29**, 698–699.

Moo-Llanes, D., Ibarra-Cerdeña, C.N., Rebollar-Téllez, E.A., Ibáñez-Bernal, S., González, C. & Ramsey, J.M. (2013). Current and future niche of North and Central American sand flies (Diptera: Psychodidae) in climate change scenarios. *PLoS Negl Trop Dis*, **7**, e2421.

Morrison, A.C., Ferro, C., Morales, A., Tesh, R.B. & Wilson, M.L. (1993). Dispersal of the sand fly *Lutzomyia longipalpis* (Diptera: Psychodidae) at an endemic focus of visceral leishmaniasis in Colombia. *Journal of medical entomology*, **30**, 427–435.

Muller, G. & Schlein, Y. (2004). Nectar and honeydew feeding of *Phlebotomus papatasi* in a focus of *Leishmania major* in Neot Hakikar oasis. *Journal of vector ecology*, **29**, 154–158.

Myers, R.A., Baum, J.K., Shepherd, T.D., Powers, S.P. & Peterson, C.H. (2007). Cascading effects of the loss of apex predatory sharks from a coastal ocean. *Science*, **315**, 1846–1850.

Myers, S.S., Gaffikin, L., Golden, C.D., Ostfeld, R.S., H. Redford, K., H. Ricketts, T., Turner, W.R. & Osofsky, S.A. (2013). Human health impacts of ecosystem alteration. *Proceedings of the National Academy of Sciences*, **110**, 18753–18760.

Nieves, E., Ora&#xe1, Luzmary, Rond&#xf3, N, Y., S&#xe1, Nchez, M., S&#xe1, Nchez, Y., Rojas, M., Rond&#xf3, N, M., Rujano, M., Gonz&#xe1, Lez, N. & Cazorla, D. (2014). Effect of Environmental Disturbance on the Population of Sandflies and *Leishmania* Transmission in an Endemic Area of Venezuela. *Journal of Tropical Medicine*, **2014**, e280629.

Ogden, N.H. & Tsao, J.I. (2009). Biodiversity and Lyme disease: Dilution or amplification? *Epidemics*, **1**, 196–206.

Oliveira, A.G., Galati, E.A.B., Fernandes, C.E., Dorval, M.E.C. & Brazil, R.P. (2008). Seasonal variation of *Lutzomyia longipalpis* (Lutz & Neiva, 1912) (Diptera: Psychodidae: Phlebotominae) in endemic area of visceral leishmaniasis, Campo Grande, state of Mato Grosso do Sul, Brazil. *Acta Tropica*, **105**, 55–61.

Orshan, L., Elbaz, S., Ben-Ari, Y., Akad, F., Afik, O., Ben-Avi, I., Dias, D., Ish-Shalom, D., Studentsky, L. & Zonstein, I. (2016). Distribution and Dispersal of *Phlebotomus papatasi* (Diptera: Psychodidae) in a Zoonotic Cutaneous Leishmaniasis Focus, the Northern Negev, Israel. *PLOS Negl Trop Dis*, **10**, e0004819.

Ostfeld, R.S. & Keesing, F. (2012). Effects of Host Diversity on Infectious Disease. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 157–182.

Peterson, A.T. & Shaw, J. (2003). *Lutzomyia* vectors for cutaneous leishmaniasis in Southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *International journal for parasitology*, **33**, 919–931.

Randolph, S.E. & Dobson, A.D.M. (2012). Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology*, **139**, 847–863.

Ready, P.D. (2013). Biology of phlebotomine sand flies as vectors of disease agents. *Annual Review of Entomology*, **58**, 227–250.

Rotureau, B., Gaborit, P., Issaly, J., Carinci, R., Fouque, F. & Carme, B. (2006). Diversity and ecology of sand flies (Diptera: Psychodidae: Phlebotominae) in coastal French Guiana. *American Journal of Tropical Medicine and Hygiene*, **75**, 62.

Schlein, Y. & Muller, G. (1995). Assessment of Plant Tissue Feeding by Sand Flies (Diptera: Psychodidae) and Mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology*, **32**, 882–887.

Shaw, J.J. & Lainson, R. (1972). Leishmaniasis in Brazil: VI. Observations on the seasonal variations of *Lutzomyia flaviscutellata* in different types of forest and its relationship to enzootic rodent Leishmaniasis (*Leishmania mexicana amazonensis*). *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **66**, 709–717.

Tesh, R.B., Chaniotis, B.N., Aronson, M.D., Johnson, K.M. & others. (1971). Natural host preferences of Panamanian phlebotomine sandflies as determined by precipitin test. *American Journal of Tropical Medicine and Hygiene*, **20**, 150–156.

de Thoisy, B., Fayad, I., Clément, L., Barrioz, S., Poirier, E. & Gond, V. (2016). Predators, Prey and Habitat Structure: Can Key Conservation Areas and Early Signs of Population Collapse Be Detected in Neotropical Forests? *PLOS ONE*, **11**, e0165362.

de Thoisy, B., Richard-Hansen, C., Goguillon, B., Joubert, P., Obstancias, J., Winterton, P. & Brosse, S. (2010). Rapid evaluation of threats to biodiversity: human footprint score and large vertebrate species responses in French Guiana. *Biodiversity and Conservation*, **19**, 1567–1584.

Wood, C.L., Lafferty, K.D., DeLeo, G., Young, H.S., Hudson, P.J. & Kuris, A.M. (2014). Does biodiversity protect humans against infectious disease? *Ecology*, **95**, 817–832.

CHAPITRE 3: LES REPAS SANGUINS DE PHLÉBOTOMES: FENÊTRE SUR LA DIVERSITÉ DE VERTÉBRÉS

RÉSUMÉ

La diversité d'hôtes est un paramètre qu'il est nécessaire d'estimer pour tester l'hypothèse de l'effet de dilution. La caractérisation de la faune de vertébrés sauvages est toutefois une tâche difficile qui demande classiquement le déploiement de méthodes observationnelles contraignantes. **Dans ce chapitre**, constitué de deux articles, nous évaluons le potentiel de l'analyse des repas sanguins de phlébotomes et de moustiques pour estimer la diversité des vertébrés sur les sites d'études.

Dans le premier article, nous évaluons la fiabilité d'un marqueur mitochondrial court (≈ 100 pb), situé dans la sous-unité ribosomale 12S, pour l'identification des mammifères amazoniens. Ce marqueur (12S-V5), conçu à l'origine pour l'identification des vertébrés par metabarcoding, présente des qualités intéressantes pour l'analyse des repas sanguins: une large couverture taxonomique des amorces PCR (les phlébotomes et moustiques peuvent se nourrir sur une grande diversité d'hôtes vertébrés) et une taille courte qui permet son amplification sur de l'ADN partiellement dégradé (dans du sang digéré). Nous constituons une base de référence moléculaire quasiment exhaustive pour les espèces présentes en Guyane, sur la base de 576 spécimens représentant 164 espèces. Nous proposons une comparaison avec les performances d'un marqueur de taille similaire situé dans le gène COI. Beaucoup de séquences de référence sont déjà disponibles pour ce gène qui contient le *barcode* d'ADN standard. Nous montrons que les deux marqueurs permettent des identifications précises (quasiment 100% d'identifications correctes dont 90% à l'espèce), mais que les sites de fixations des amorces PCR sont beaucoup plus conservés dans le cas du marqueur ribosomal, ce qui constitue un avantage considérable.

Dans le deuxième article, nous appliquons la méthode décrite dans le premier article pour l'identification des repas sanguins de phlébotomes et de moustiques piégés dans la régions de Saint-Georges. L'analyse d'environ 200 spécimens gorgés collectés sur 3 sites a permis la détection de 22 espèces de vertébrés incluant des mammifères arboricoles et terrestres de tailles variées ainsi que des oiseaux, des lézards et des amphibiens. La comparaison des résultats obtenus entre les différents sites indique une diversité de vertébrés moindre sur les sites sujets à des perturbations anthropiques plus importantes. Ceci suggère que l'analyse des repas sanguins de diptères peut constituer une alternative intéressante pour la description et la comparaison des communautés de vertébrés. Par ailleurs, notre étude a permis de générer des données inédites concernant les préférences trophiques des

phlébotomes, ce qui représente une contribution importante pour l'étude de la transmission des leishmanioses.

ARTICLE 1:
EVALUATION OF SHORT MITOCHONDRIAL METABARCODES FOR THE IDENTIFICATION OF AMAZONIAN MAMMALS

Arthur Kocher[1,2], Benoit de Thoisy[3,4], François Catzeflis[5], Mailis Huguin[3,4], Sophie Valière[6], Lucie Zinger[1], Anne-Laure Bañuls[2] and Jérôme Murienne[1*].

Published in *Methods in Ecology and Evolution*
(doi: 10.1111/2041-210X.12729)

[*1]CNRS, Université Toulouse III Paul Sabatier, ENFA; UMR5174 EDB (Laboratoire Evolution et Diversité Biologique), Toulouse, France; [2]UMR MIVEGEC (IRD 224 – CNRS 5290 – Université de Montpellier), 911 Avenue Agropolis, F34394 Montpellier, France; [3]Institut Pasteur de la Guyane, 23 avenue Pasteur, 97300 Cayenne, French Guiana; [4]Association Kwata, 16 avenue Pasteur, 97300 Cayenne, French Guiana; [5]Institut des Sciences de l'Evolution, Case Courrier 064, CNRS UMR-5554, Université Montpellier-2, Place E. Bataillon, F-34095 Montpellier, France; [6]GeT–PlaGe, Genotoul, INRA Auzeville, 31326 Castanet-Tolosan, France.

ABSTRACT

DNA barcoding and metabarcoding are increasingly used as alternatives to traditional morphological identifications. For animals, the standard barcode is a ~658-bp portion of the COI gene, for which reference libraries now cover a large proportion of described mammal species. Unfortunately, because its sequence is too long and does not contain highly conserved primer binding sites, this marker is not adapted for metabarcoding. Although alternative metabarcodes have been developed, their performance is generally seldom assessed.

Here, we evaluate the reliability of a short metabarcode located in the mitochondrial 12S ribosomal RNA for identifications of Amazonian mammals. We (i) constitute a nearly exhaustive reference library for species found in French Guiana, (ii) assess the taxonomic resolution of the marker and validate its use with dipteran blood meal analyses, (iii) assess the conservation of the primer binding sites and (iv) compare its theoretical performances with that of a newly designed metabarcode located within the standard COI barcode. 576 specimens representing 164 species were gathered and sequenced. We show that the 12S marker allows remarkably accurate taxonomic assignations despite its very short size, and that primer binding sites are highly conserved, which is important to avoid PCR amplification bias potentially leading to detection failure. Additionally, our results stress that identifications should only be considered at the generic level when they are based on incomplete reference libraries, even when a stringent similarity cut-off is used. A new short COI metabarcode was designed based on 569 reference sequences of mammals retrieved on BOLD. Our results clearly show that, while both markers provide similar taxonomic resolution, much higher rates of primer mismatches are found with COI.

Besides demonstrating the accuracy of the short 12S marker for the identification of Amazonian mammals and providing a reliable molecular reference database, this study emphasize that the accuracy of taxonomic assignations highly depends on the comprehensiveness of the reference library and that great caution should be taken for interpreting metabarcoding results based on scarce reference libraries. The comparison with a short COI metabarcode also provides novel evidence in support for the use of ribosomal markers in metabarcoding studies.

INTRODUCTION

The accurate identification of species is an essential component in most of empirical ecological studies. Traditional methods based on morphological features are time consuming and often rely on taxonomic expertise that is increasingly lacking. In addition, morphological identifications may require whole specimens, which can be particularly difficult to obtain for mammals due to practical, ethical or legal reasons. DNA-based identification methods have been increasingly used as an efficient alternative over the last decades. Today, one of the most used techniques is DNA barcoding (Hebert *et al.* 2003), which uses the sequence from a short standard fragment of the genome for the taxonomic assignment of a specimen. More recently, high-throughput sequencing has allowed the extension of DNA barcoding for the identification of multiple species in a single sample (Taberlet *et al.* 2012). This approach, referred to as metabarcoding, allows the simultaneous identifications of multiple specimens from a single bulk-DNA extraction (Yu *et al.* 2012; Kocher *et al.* 2016). In addition, it has the great advantage to be applicable on degraded DNA present in the environment such as soil (Andersen *et al.* 2012) or water (Ficetola *et al.* 2008; Valentini *et al.* 2016). Finally, it constitutes a great tool to study trophic interactions through the analyses of gut content (Coghlan *et al.* 2013) or faeces (Kartzinel *et al.* 2015).

The prerequisites of these methods are the choice of appropriate DNA markers, the design of corresponding PCR primers and the constitution of reliable reference sequences libraries. For DNA barcoding, the Consortium for the Barcode of Life (CBOL, http://www.barcodeoflife.org/) handled these issues, by providing standardized laboratory protocols and curated reference libraries linked to voucher specimens. For animals, the current standard barcode is a ~658-bp portion of the mitochondrial cytochrome oxidase 1 subunit (COI), and the Barcode of Life Database comprises reference sequences for more than 174,000 animal species to date (BOLD, http://www.boldsystems.org/, accessed in Sep. 2016). Unfortunately, this marker is not the best choice when it comes to metabarcoding (Deagle *et al.* 2014). First, the fragment is too long regarding the limitations of the current sequencing platforms (typically an Illumina Miseq; Illumina, Inc., San Diego, CA, USA). This can be regarded as a rather technical issue that might be overcome in a near future with the rapid improvement of the sequencing technologies. However, the size of the targeted fragment is also critical when dealing with degraded DNA, as typically found in the environment or in biological samples such as faeces or gut content. Secondly, it is virtually impossible to find perfectly conserved primer binding sites within this coding gene because of high mutation

rate at the third codon position (Deagle *et al.* 2014). This may not be a problem for barcoding single specimens, because a few primer-template mismatches will not impede PCR amplification. On the contrary, small variations in the number and position of primer-template mismatches can lead to significant amplification bias or even detection failure when mixtures of DNA are amplified for metabarcoding (Bru *et al.* 2008; Taberlet *et al.* 2012). In order to find suitable metabarcodes and their associated primers, specific softwares have been developed (notably "ecoPrimers", (Riaz *et al.* 2011)), that seek to minimize amplification bias while maximizing the divergence between taxa. Most animal metabarcoding markers developed using this approach are located within the mitochondrial ribosomal RNA genes (Riaz *et al.* 2011; Clarke *et al.* 2014; Deagle *et al.* 2014). Indeed, because of the secondary structure of their RNA products, these genes exhibit a mosaic pattern of variation with highly conserved regions (within stems) in which primers can be designed, adjacent to variable regions (within loops) that allow interspecific discrimination. The existence of a standard marker for DNA barcoding has allowed the constitution of a collaborative and taxonomically comprehensive reference library. On the contrary, there is no real consensus on the choice of metabarcoding markers (except for bacteria and fungi), leading to scarce reference libraries (Pompanon & Samadi 2015).

The 12S-V5 marker is a *c.* 100 base pairs (bp) portion of the mitochondrial 12S ribosomal RNA gene (12S rRNA) (Riaz *et al.* 2011). Based on the sequences available in public databases, it was shown to gather good properties for metabarcoding of vertebrates. However, a comprehensive taxonomic sampling is necessary to precisely assess the taxonomic resolution of a DNA marker (Meyer & Paulay 2005). In this study, we assess the reliability of the 12S-V5 markers for metabarcoding of Amazonian mammals. We (i) constitute a nearly exhaustive reference library for the species found in French Guiana, (ii) assess the variability at the 12S-V5 primer binding sites, (iii) evaluate the taxonomic resolution of the marker and further validate its use with dipteran blood meal analyses and (iv) compare its theoretical performances with that of a newly designed metabarcode located within the classical COI barcode.

MATERIAL AND METHOD

Sampling

French Guiana and its >90% of well-preserved Amazonian rainforest cover has been the study site of intense ecological and taxonomic research (see for instance the research

undertaken under the frame of the labex CEBA; http://www.labex-ceba.fr/en/). The mammalian fauna of French Guiana is well characterized, and is representative of a larger part of the northern Amazon region (Lim *et al.* 2012; Catzeflis 2015). Our aim was to generate a first comprehensive DNA library for the mammals of French Guiana that can be used directly for metabarcoding studies in this region, and that may be further completed for studies in other Amazonian locations. Tissue samples of mammals from French Guiana were gathered from field sampling, museum collections, hunting or road-killed specimens and biopsies of captured animals (see Supporting Information for details). The taxonomic identifications were based on external and/or craniodental morphology, and confirmed by COI barcoding for the vast majority of the specimens [following classical procedures (Borisenko *et al.* 2008) and the primers C_VF1di / C_VR1LRt1 or LCO1490 / HCO2198 recommended by the Barcode of Life Project (www.boldsystems.org)].

Reference library

Our aim was to build a reference sequence library based on a marker that was previously developed for identifications of vertebrates through metabarcoding (12S-V5, (Riaz *et al.* 2011)). The constitution of reference libraries with previously designed PCR primers leads to the loss of all the information regarding the primer binding sites. This impedes the possibility to further improve the primers for specific purposes and to predict potential amplification bias (Bru *et al.* 2008). To overcome this issue, we designed a set of primers to amplify a region that contains the complete 12S-V5 fragment including primer binding sites (see Fig. 1). We used blastn 2.2.29+ (Camacho *et al.* 2009) on Genbank (release 197) with the query being a *c*. 700 bp portion of *Rattus rattus* 12S rRNA (GenBank accession: NC_012374.1) containing the fragment amplified by the 12S-V5 primers and 300 bp flanking each end. We selected all the matching sequences of mammals presenting at least 95% query coverage and kept one sequence per species. The resulting database contained sequences for 1557 mammal species representing 26 orders and was used to design new primers with the `ecoPrimers` program (Riaz *et al.* 2011). These primers (Mam12S-340-F, 5'-CCACCGCGGTCATACGATT-3'; Mam12S-340-R, 5'- GATGGCGGTATATAGACTG-3') had a maximum of two mismatches with 98.4% of the species represented in the database. They amplify a fragment of 302 to 350 bp that contains the full 12S-V5 marker and can be sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA, USA).

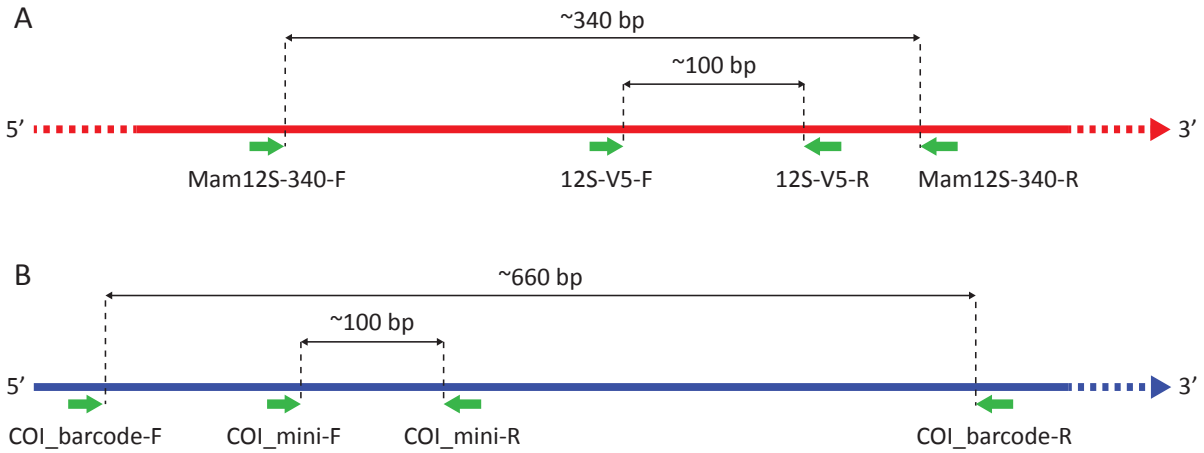


Figure 1: (A) Relative positions of the Mam12S-340 and 12S-V5 primers binding sites on the 12S mt rRNA gene. PCR amplifications were performed using Mam12S-340 primers in order to generate a reference library for the 12S-V5 metabarcode while keeping the information regarding the primers binding sites. **(B)** Relative positions of the standard COI barcode and the newly designed COI_minimam primers on the COI gene. The COI reference library was constituted with full standard barcodes allowing to investigate COI_minimam primers binding sites.

DNA amplification and sequencing

DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, USA). PCR amplification was performed in 25 µl mixtures containing 2 µl of DNA template, 0.2 µl of AmpliTaq Gold® (5U/ µl; Applied Biosystems, Foster City, CA, USA), 2.5 µl 10X PCR buffer (provided with Amplitaq Gold®), 0.5 µl dNTPs (2.5 mM each, Promega, Madison, WI, USA), 1 µl of each primer (10 µM), 0.25 bovine serum albumin (10mg/ml, Promega), 2.5 µl $MgCl_2$ (25 mM, Applied Biosystems) and nuclease-free water (Promega). The PCR mixture was denatured at 95°C (10 min) and followed by 35 cycles of 30s at 95 °C, 30s at 50°C and 30s at 72 °C, completed at 72 °C (10 min). Tags of eight base pairs with at least five differences between them were added at the 5' end of each primer to enable the sequencing of the multiple PCR products in a single sequencing run (Binladen *et al.* 2007).

PCR products were pooled and sent for library construction and sequencing to the GeT-PlaGe core facilities of Genotoul (Toulouse, France). Samples were diluted in ultrapure water. A volume of 130 µl containing 3 µg of DNA was purified using the HighPrep PCR system (Magbio Genomics, Gaithersburg, MD, USA) and used for library construction with the Illumina NEXTflex PCR-Free DNA sequencing kit following the instructions of the supplier (Bioo Scientific corp., Austin, TX, USA). Purified fragments were end-repaired, A-tailed and ligated to sequencing indexed adapters. The quality of the library was controlled using the Fragment Analyzer (Advanced Analytical, Ames, IA, USA) and quantified by qPCR with the Library Quantification Kit - Illumina Genome Analyzer-SYBR Fast Universal (CliniSciences, Nanterre, France). The library was loaded onto the Illumina MiSeq cartridge according to the

manufacturer instructions. The quality of the run was checked internally using PhiX. Quality filtering was performed by the Consensus Assessment of Sequence and Variation (CASAVA) pipeline. The sequencing data was stored on the NG6 platform (Mariette *et al.* 2012) and all computations were performed on the computer cluster of the Genotoul bioinformatic platform (Toulouse, France).

Bioinformatics

Sequence reads were analysed using the `OBITOOLS` package (Boyer *et al.* 2016). Pair-end reads were aligned and merged, taking into account the phred quality scores for consensus construction and alignment score computation. The reads were then assigned to their corresponding sample based on the tagged primer sequences with two mismatches allowed. Low quality reads (alignment scores < 50, containing Ns or shorter than 50bp) were removed. Reads were then dereplicated while keeping the coverage information (number of reads merged). For each sample, the majority sequence was considered as the genuine most abundant sequence in the specimen and kept for the reference library. The script used for these bioinformatic steps are available in the Supporting Information. The library was further completed by 12S rRNA sequences extracted from complete mitogenomes of Xenarthra (Gibb *et al*. 2016) and Chiroptera (Botero-Castro *et al*., unpublished).

Metabarcode evaluation

Primer-template mismatches were checked by mapping the 12S-V5 primers on the resulting sequences (see Fig. 1; 12S-V5-F: TAGAACAGGCTCCTCTAG; 12S-V5-R: TTAGATACCCCACTATGC), using Geneious 6.0.6 Pro (Biomatters, Auckland, New Zealand). The region corresponding to the 12S-V5 metabarcode was then extracted from each sequence for the following analyses.

Most studies that validate the reliability of DNA barcoding for molecular identifications provide statistics based on K2P genetic distances (Kimura 1980) computed from a multiple sequence alignment of the reference sequences. Uncorrected distances have been judged more appropriate for studying the success of distance-based identification techniques (Srivathsan & Meier 2012). Here, we used genetic distances as they are computed by the `ecotag` program (included in the `OBITOOLS`) for taxonomic assignments (uncorrected distances based on pairwise alignments of the sequences) to generate a neighbour-joining tree using the R package 'ape' (Saitou & Nei 1987; Paradis *et al.* 2004; Team 2014) and to compute distance statistics using the R package 'spider' (Brown *et al.* 2012).

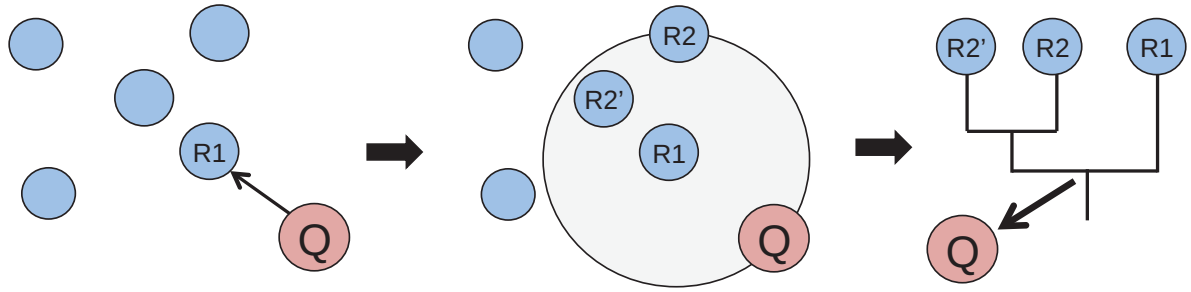


Figure 2: Schematization of a taxonomic assignation as performed by ecotag: (i) search of the reference sequences (R1s) that have the highest similarity with the query sequence (Q), (ii) search for all other reference sequences (R2s) whose similarity with the primary reference sequences is equal or higher than the similarity between the primary reference sequences and the query sequences, (iii) assignment of the query sequence to the most recent clade containing all R1s and R2s.

In order to assess the reliability of the metabarcode for species identifications, we performed the taxonomic assignment of each specimen using the `ecotag` program with all other sequences as reference library. Resulting assignations were then compared to the genuine identities of the specimens. `Ecotag` first searches for the reference sequence(s) showing the highest similarity with the query sequence (primary reference sequence(s); see Fig. 2). Then it looks for all other reference sequences whose similarity with the primary reference sequence(s) is equal or higher than the similarity between the primary reference sequence(s) and the query sequence (secondary reference sequence(s)). Finally, it assigns the query sequence to the most recent common ancestor of the primary and secondary reference sequences. This procedure is similar in essence to the lowest common ancestor algorithms implemented in the MG-RAST server (Meyer *et al.* 2008) and the MEGAN program (Huson *et al.* 2007) for the assignment of metagenomic reads. It allows to deal with ambiguous identifications, which can be due to the fact that several taxa are poorly distinguishable, or that the DNA library does not contain a close reference for the query. Taxonomic assignations were discarded if the closest match exhibited less than 97% similarity. To assess the effect of potential taxonomic gaps in the reference database, we then tested the taxonomic assignations of each sequence after removing all conspecifics. To further validate the applicability of the 12S-V5 marker for mammal species identification in field studies, we analysed blood meals of hematophagous dipteran collected in forest sites in French Guiana (sand flies and mosquitoes, see Supporting information for details on sampling and laboratory protocols). Indeed, metabarcodes provide good properties (small size and wide taxonomic coverage) for such application, because arthropod blood meals may contain low quantities of degraded DNA from a diverse array of vertebrate species.

Comparison with COI

Currently, COI is rarely used for metabarcoding because of previously explained reasons (see introduction). In particular, no satisfying COI metabarcode has been developed for mammals. Therefore, to allow relevant comparison, we designed new PCR primers to amplify a short fragment located within the classical COI barcode. All the sequences of mammal species found in French Guiana were retrieved from BOLD, and a maximum of five sequences per species were kept. PCR primers were designed using `ecoPrimers` in the same way it was done for the 12S-V5 primers (*i.e*. 18 bp-long, in order to amplify a fragment of *c.* 100 bp, and to maximise taxonomic coverage and resolution). The selected primers were compared to their target sites on the reference sequences in order to compute mismatch statistics. The theoretical amplified fragment was then extracted from references sequences to evaluate its taxonomic resolution in the same way it has been done for the 12S-V5 marker.

RESULTS

Reference library

Sequences were obtained for 576 specimens representing 164 species, including *Uroderma* cf. *magnirostrum* (although *U. magnirostrum* has not been officially reported in French Guiana, and no voucher specimen is available, it is not unlikely that this taxon occurs and the samples used in this study have been identified as *U. magnirostrum* using classical COI barcoding). Hence, 82.7% (163/197; (Catzeflis 2015)) of the mammal species recorded in French Guiana were included in the database, whereas 116 over 126 genera and all the 32 families were represented. All the sequences were deposited on Genbank (accessions: KX381203-KX381784). On average, 3.5 specimens per species were sequenced with 118 species (72%) being represented by at least 3 specimens and 28 (17%) represented by only one. The sequence length of the Mam12S-340 fragment ranged from 334 to 345 bp whereas the length of the 12S-V5 fragment ranged from 96 to 103 bp**.**

Metabarcode evaluation

The inspection of the 12S-V5 primer binding sites revealed that the forward primer could be slightly improved by degenerating the 5' end (12S-V5-F': YAGAACAGGCTCCTCTAG). 95.0%, 2.8% and 2.2 % of the sequences had respectively 0,1 and 2 mismatches with the forward primer (mean number of mismatches = 0.072; degenerated version), whereas 97.6% and 2.4 % had respectively 0 and 1 mismatch with the

reverse primers (mean number of mismatches = 0.024). No sequence had mismatches with both primers. No sequence had mismatch toward the 3' end on the forward primer (within the first eight positions). One mismatch was found at the second position from the 3' end of the reverse primer in *Tonatia saurophila* and *Rhynophylla pumilio*, which may hamper amplification for these species.

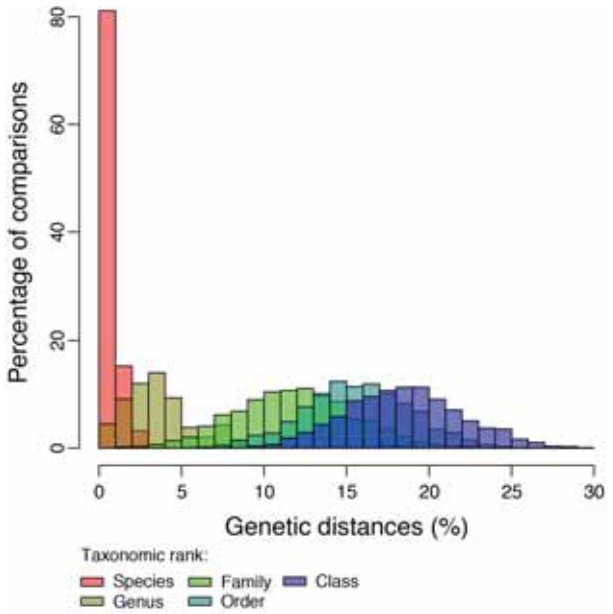


Figure 3: Distribution of genetic distance at various taxonomic resolutions. Only the distances between specimens belonging to (i) the same taxon at a given rank and to (ii) distinct taxa at the inferior rank are considered.

In the 12S-V5 metabarcode, 34.3% of sites were identical in all sequences and the mean pairwise identity was of 80.2%. The distribution of genetic distances at the specific, generic, familial and ordinal levels shows significant overlap between each consecutive taxonomic rank (Fig. 3). In particular, intraspecific distances range from 0.0% to 5.1% for a mean of 0.5%, while interspecific distances within the same genus range from 0.0% to 19.6% for a mean of 3.5%. The neighbour-joining tree is presented in Fig. 4. Some closely related species were poorly or not distinguishable based on the marker, mostly in Chiroptera and in Felidae: *Pteronotus rubiginosus* and *Pteronotus* cf. sp.3, *Molossus molossus* and *Molossus rufus, Carollia brevicauda* and *Carollia perspicillata*, *Eptesicus furinalis* and *Eptesicus chiriquinus*, *Artibeus planirostris* and *Dermanura gnoma, Oecomys rex* and *Oecomys* sp. 1, *Puma concolor* and *Puma yagouaroundi* and *Leopardus wiedii* and *Leopardus pardalis*.

When considering the species represented by more than one specimen (548 sequences), 90.0% and 9.5% of the assignations were made at the species and genus level, respectively (Tab. 1). 99.6% of these assignations were correct. Only two errors were found: a sequence of

Leopardus wiedii was assigned to *Leopardus pardalis* and a sequence of *Dermanura gnoma* was assigned to the genus *Artibeus.* Most of the assignations made at the genus level were found in Chiroptera (genera *Carollia, Eptesicus*, *Molossus* and *Pteronotus*). Three sequences (two *Glyphonycteris sylvestris* and one *Makalata didelphoides*) were not identified due to the absence of a close match in the reference database (>97% similarity). When all conspecifics were removed from the references before taxonomic assignment, 65.2% of the sequences were not identified because they did not find a close match, as it was expected. However, 23.6% of the sequences were falsely assigned at the specific level. 10.5% of the sequences were assigned at the generic rank with an error rate of 41%, corresponding to confusion between the genera *Artibeus* and *Dermanura*, *Didelphis* and *Philander* and *Puma* and *Panthera*. Finally, four sequences were assigned at the family level, all correctly. The overall proportion of specimens that were assigned to a wrong taxon was 26.9%. If identifications made at the specific level were only considered at the generic level, they would have been correct in 93.9% of cases, and the overall proportion of misidentified specimens would have been lowered to 4.7%.

Table 1: Results of taxonomic assignations using ecotag: each specimen was considered as unknown while all other sequences were used as references. Percentages of identifications that were made at each taxonomic rank are indicated, as well as the percentage of specimens that could not be identified because they did not find a close match in the reference library.

Metabarcode	Reference library	Proportion of identifications made at given taxonomic rank					Overall error rate†
		Species	Genus	Family	Order	Not identified	
12S-V5	Comprehensive*	90.0	9.5	0.0	0.0	0.5	0.4
	No conspecific§	23.6	10.5	0.7	0.0	65.2	26.9
COI_minibarcode	Comprehensive*	87.7	4.7	0.0	0.0	7.6	1.1
	No conspecific§	10.1	0.0	0.0	0.0	89.9	10.1

*All specimens have at least one conspecific to match against in the reference library

§To assess the effect of potential gaps in the database, we removed every conspecific from the references before taxonomic assignations

‡Percentage of specimens that were wrongly identified

Thirty blood-fed specimens, including four mosquitoes and 26 sand flies, were collected in French Guiana. Amplification and sequencing of the 12S-V5 marker was successful for 27 of the specimens, and allowed the identification of eight mammal species belonging to five distinct orders: Primate, Didelphimorphia, Rodentia, Carnivora and Xenarthra (see Supporting information for details).

Figure 4: Neighbour-joining tree based on raw distances computed from pairwise alignments of the 12S-V5 metabarcode. Numbers in brackets indicate the number of specimens sequenced per species. Species names are coloured by order.

Comparison with COI

A total of 569 COI sequences representing 138 mammal species found in French Guiana were retrieved from BOLD. The newly designed PCR primers amplified a 102 pb-long fragment, and had similar theoretical melting temperatures (COI_minimam_F: 5'-CCCATGCATTCGTAATAA-3'; COI_minimam_R: 5'-GTAGAAGTCAGAAGCTTA-3'). The number of mismatches per reference sequence ranged from 0 to 5 for both primers for a mean of 2.13 and 2.2 for primer F and R respectively. No reference sequence had zero mismatch with both primers. To visualize mismatching patterns of 12S and COI primers, an alignment of each primer pair with reference sequences of mammals from various orders is

presented in Fig. 5. Assessment of the COI_minimam taxonomic resolution revealed that with a comprehensive reference database, 87.7% and 4.7% of sequences would have been assigned at the specific and generic level respectively, while, 7.6% would have remained unidentified, for an overall error rate of 1.1% (Tab. 1). With an incomplete reference database, 10.1% of the sequences would have been wrongly identified at the specific level, while the all others would have remained unidentified (*i.e.* an overall error rate of 10.1%).

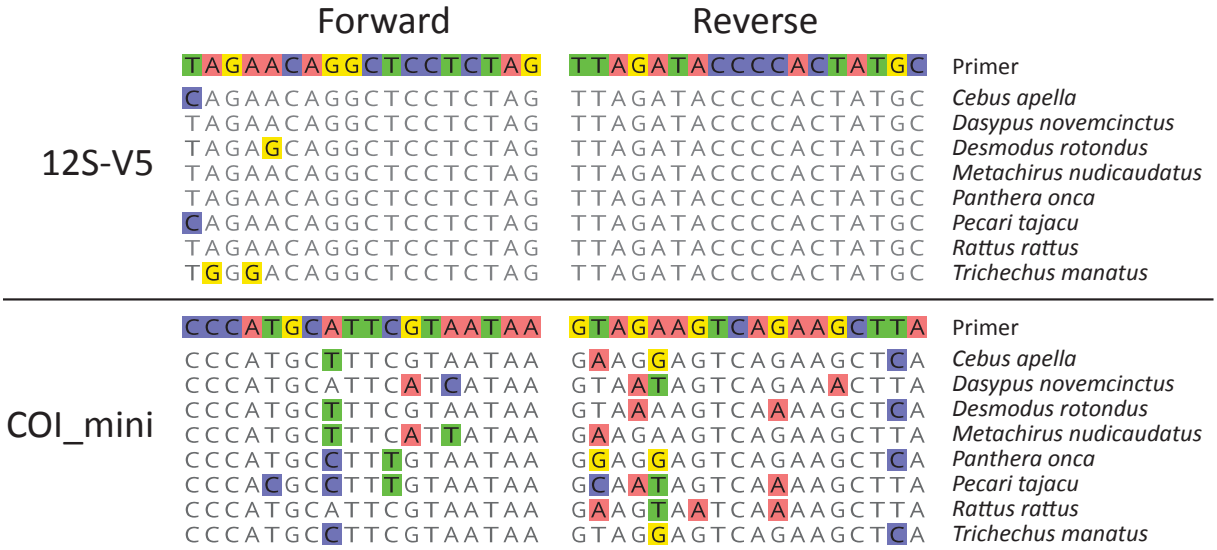


Figure 5: Alignment of 12S-V5 and COI_minimam primers with reference sequences of mammals belonging to distinct orders. Primer mismatches are highlighted in the reference sequences.

DISCUSSION

By enabling the identification of species from degraded DNA contained in environmental samples, DNA metabarcoding has opened great avenues for the study of vertebrates' communities. It has already been proved successful in various applications such as the characterisation of the present or ancient terrestrial fauna from the soil (Andersen *et al.* 2012; Giguet-Covex *et al.* 2014), aquatic communities from water (Valentini *et al.* 2016) or feeding behaviours from faeces (De Barba *et al.* 2014). Because the standard COI barcode is not adapted for metabarcoding (Deagle *et al.* 2014), these studies relied on the development of other markers.

First, we show that the 12S-V5 primer binding sites are extremely conserved among the mammal species included in our database. Secondly, we show that the marker allows reliable and precise identifications of mammals, which was further highlighted by the successful analysis of dipteran blood meals. Using a similarity cut-off of 97%, almost all specimens could be correctly identified and more than 90% of these identifications were made at the

specific level, while the others were made at the generic level. Only two sequences were assigned to a wrong taxon (false positive errors), while three specimens were not identified because their closest match did not reach the similarity cut-off (false negative errors). This level of accuracy is remarkable regarding that the marker is only 100-bp-long and can be amplified with the same PCR primers virtually in all vertebrates (Riaz *et al.* 2011). Our results emphasize that the quality of taxonomic assignations is largely dependent on the comprehensiveness of the reference database. Indeed, when all conspecific sequences were removed prior to taxonomic assignations, almost 30% of the specimens were assigned to a wrong taxon. This error rate is largely dependent on the choice of the similarity cut-off. A more stringent (higher) cut-off would have resulted in a lower false positive error rate, but also in a higher false negative error rate. The wide overlap observed between intraspecific and interspecific genetic distances, and thus, the absence of a clear barcoding gap precludes the existence of a perfect similarity cut-off. The choice of an optimum was not in the scope of this work. However, we have shown that bringing the taxonomic assignations to the generic rank restores the reliability of the identifications when the species of the query is not represented in the reference library. Therefore, besides using a stringent similarity threshold, we recommend to consider taxonomic assignations only at the generic level when using the 12S-V5 metabarcode with an incomplete reference library. The importance of the comprehensiveness of reference libraries to avoid erroneous identifications and the difficulty to define similarity cut-off due to overlapping intra and inter-specific genetic distances has already been highlighted for classical COI barcoding (Meyer *et al.* 2008; Puillandre *et al.* 2009). This should be even more significant for shorter and less discriminant metabarcodes. Numerous studies provide thoroughly sampled COI reference libraries together with an evaluation of barcoding in specific animal groups. On the contrary, DNA metabarcoding studies usually rely on public databases, and the current literature mainly focuses on biomolecular and bioinformatic issues, such as the management of PCR and sequencing artefacts, rather than providing evaluation of metabarcodes accuracy based on comprehensive reference libraries. This is understandable because metabarcodes frequently target very wide taxonomic ranges or taxa in which a large proportion of species are unknown. In addition, it exists no real standard for metabarcoding markers (at the exception of bacteria and fungi), which impedes the creation of a single collaborative and well-curated reference database as the BOLD for barcoding (Pompanon & Samadi 2015). Nevertheless, we argue that these limitations should be considered seriously when interpreting metabarcoding results.

Finally, we compared the theoretical performances of the 12S-V5 marker with that of a newly designed COI metabarcode. The COI reference database retrieved on BOLD contained similar number of sequences and taxonomic coverage than the 12S reference database constituted in this study, which allowed relevant comparisons. The COI_minimam was selected using the same procedure than for the 12S-V5 marker. However, it is noteworthy that 12S-V5 primers were designed to amplify DNA of all vertebrates, whereas the COI_minimam was specifically directed to Amazonian mammals, which represent a rather conservative approach for our comparison. We show that while both metabarcodes provide comparable taxonomic resolution, the COI_minimam primers present high rates of primer-binding site mismatches (Fig. 5), constituting a serious disadvantage for metabarcoding studies.

Besides providing reliable and comprehensive molecular data for the identification of mammals in French Guiana and more largely for the northern Amazon region, our study emphasize that great caution should be taken regarding metabarcoding results based on scarce reference libraries and that molecular identifications should be trust only using a stringent similarity threshold and at appropriate taxonomic ranks. Our results also provide novel evidence in support for the use of ribosomal markers in metabarcoding studies.

ACKNOWLEDGMENTS

This work was supported by “Investissement d’Avenir” grants managed by Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; TULIP: ANR-10-LABX-41, ANR-11-IDEX-0002-02) as well as project METABAR (ANR-11-BSV7-0020). Many samples used in this study are part of the JAGUARS collection, supported by DEAL Guyane, Collectivité Territoriale de Guyane, and European ERDF funds. Most of COI barcoding was done within the Guyamazon II project project “Biodiversidade e zoogeografia de pequenos mamíferos no escudo das Guianas “ which was funded by Ambassade de France in Brazil, IRD / AIRD (France), the Collectivité Territoriale de Guyane, CIRAD, FAPEAM, FAPEMA, and FAPEAP. We would like to thank Frederic Delsuc and Fidel Botero-Castro (Institut des Sciences de l'Evolution de Montpellier) for providing sequences of Xenarthra and Chiroptera, as well as Pierre Taberlet and Eric Coissac for fruitful discussions about metabarcoding.

DATA ACCESSIBILITY

All sequences were deposited in GenBank (accessions: KX381203-KX381784)

AUTHOR CONTRIBUTIONS

J.M., AL.B. and A.K. designed the study. B.d.T. and F.C. provided the samples. A.K., M.H., B.d.T. and S.V. performed the laboratory work. A.K. analysed of the data. A.K. wrote the manuscript and all authors contributed in its improvement.

REFERENCES

Andersen, K., Bird, K.L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjær, K.H., Orlando, L., Gilbert, M.T.P. & Willerslev, E. (2012). Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, **21**, 1966–1979.

Binladen, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R. & Willerslev, E. (2007). The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLoS ONE*, **2**.

Borisenko, A.V., Lim, B.K., Ivanova, N.V., Hanner, R.H. & Hebert, P.D.N. (2008). DNA barcoding in surveys of small mammal communities: a field study in Suriname. *Molecular Ecology Resources*, **8**, 471–479.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. & Coissac, E. (2016). obitools: a unix-inspired software package for DNA metabarcoding. *Molecular ecology resources*, **16**, 176–182.

Brown, S.D., Collins, R.A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C.J. & Cruickshank, R.H. (2012). Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, **12**, 562–565.

Bru, D., Martin-Laurent, F. & Philippot, L. (2008). Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. *Applied and environmental microbiology*, **74**, 1660–1663.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, **10**, 421.

Catzeflis, F. (2015). Liste des Mammifères de Guyane, pp. 226-239 in *Nature guyanaise: 50 ans de progrès et de souvenirs* (L. Sanite, ed.). Editions Orphie, Saint-Denis de la Réunion.

Clarke, L.J., Soubrier, J., Weyrich, L.S. & Cooper, A. (2014). Environmental metabarcodes for insects: *in silico* PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, **14**,1160-1170.

Coghlan, M.L., White, N.E., Murray, D.C., Houston, J., Rutherford, W., Bellgard, M.I., Haile, J. & Bunce, M. (2013). Metabarcoding avian diets at airports: implications for birdstrike hazard management planning. *Investigative Genetics*, **4**, 27.

De Barba, M., Adams, J.R., Goldberg, C.S., Stansbury, C.R., Arias, D., Cisneros, R. & Waits, L.P. (2014). Molecular species identification for multiple carnivores. *Conservation Genetics Resources*, **6**, 821–824.

Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F. & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, **10**, 20140562.

Ficetola, G.F., Miaud, C., Pompanon, F. & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology letters*, **4**, 423–425.

Gibb, G.C., Condamine, F.L., Kuch, M., Enk, J., Moraes-Barros, N., Superina, M., Poinar, H.N. & Delsuc, F. (2016). Shotgun Mitogenomics Provides a Reference Phylogenetic Framework and Timescale for Living Xenarthrans. *Molecular Biology and Evolution*, **33**, 621–642.

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gielly, L., Domaizon, I., Coissac, E., David, F., Choler, P., Poulenard, J. & Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications*, **5**.

Hebert, P.D., Cywinska, A., Ball, S.L. & others. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, **270**, 313–321.

Huson, D.H., Auch, A.F., Qi, J. & Schuster, S.C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, **17**, 377–386.

Kartzinel, T.R., Chen, P.A., Coverdale, T.C., Erickson, D.L., Kress, W.J., Kuzmina, M.L., Rubenstein, D.I., Wang, W. & Pringle, R.M. (2015). DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. *Proceedings of the National Academy of Sciences*, **112**, 8019–8024.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, **16**, 111–120.

Kocher, A., Gantier, J.-C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., Dusfour, I., Girod, R., Bañuls, A.-L. & Murienne, J. (2016). Vector soup: high-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources*, in press.

Lim, B.K., Patterson, B.D. & Costa, L.P. (2012). Biogeography of mammals from the Guianas of South America. *Bones, clones and biomes: the history and geography of recent Neotropical mammals*. University of Chicago Press, Chicago, Illinois, 230–258.

Mariette, J., Escudié, F., Allias, N., Salin, G., Noirot, C., Thomas, S. & Klopp, C. (2012). NG6: Integrated next generation sequencing storage and processing environment. *BMC genomics*, **13**, 462.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. & Edwards, R.A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*, **9**, 386.

Meyer, C.P. & Paulay, G. (2005). DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLOS Biol*, **3**, e422.

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.

Pompanon, F. & Samadi, S. (2015). Next generation sequencing for characterizing biodiversity: promises and challenges. *Genetica*, **143**, 133-138.

Puillandre, N., Strong, E.E., Bouchet, P., BOISSELIER, M.-C., Couloux, A. & Samadi, S. (2009). Identifying gastropod spawn from DNA barcodes: possible but not yet practicable. *Molecular Ecology Resources*, **9**, 1311–1321.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P. & Coissac, E. (2011). ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, 1-11.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, **4**, 406–425.

Srivathsan, A. & Meier, R. (2012). On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics*, **28**, 190–194.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, **21**, 2045–2050.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.-M., Peroux, T., Crivelli, A.J., Olivier, A., Acqueberge, M., Le Brun, M., Møller, P.R., Willerslev, E. & Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, **25**, 929–942.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C. & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**, 613–623.

APPENDICES

Appendix 1: List of the specimens used in this study along with detail information

ID	Taxon	Voucher ID	Voucher localisation*	Sampling country	COI done	Genbank accession
B1124	*Lasiurus blossevillii*	/	/	French Guiana	yes	KX381203
C1000	*Cynomops paranus*	/	/	French Guiana	yes	KX381204
C1001	*Cynomops paranus*	/	/	French Guiana	yes	KX381205
C1005	*Artibeus planirostris*	/	/	French Guiana	yes	KX381206
C1018	*Chrotopterus auritus*	/	/	French Guiana	yes	KX381207
C1021	*Lonchorhina inusitata*	/	/	French Guiana	no	KX381208
C1022	*Vampyrodes caraccioli*	/	/	French Guiana	yes	KX381209
C1023	*Vampyrum spectrum*	/	/	French Guiana	yes	KX381210
C1025	*Chiroderma villosum*	/	/	French Guiana	yes	KX381211
C1026	*Artibeus lituratus*	/	/	French Guiana	yes	KX381212
C1027	*Artibeus lituratus*	/	/	French Guiana	yes	KX381213
C1031	*Lophostoma brasiliense*	/	/	French Guiana	no	KX381214
C1100	*Uroderma cf. magnirostrum*	/	/	French Guiana	yes	KX381215
C1102	*Uroderma cf. magnirostrum*	/	/	French Guiana	yes	KX381216
C1123	*Platyrrhinus brachycephalus*	/	/	French Guiana	yes	KX381217
C1142	*Cormura brevirostris*	/	/	French Guiana	no	KX381218
C1179	*Saccopteryx bilineata*	/	/	French Guiana	yes	KX381219
C1182	*Sturnira tildae*	/	/	French Guiana	yes	KX381220
C1184	*Dermanura cinerea*	/	/	French Guiana	yes	KX381221
C1185	*Dermanura cinerea*	/	/	French Guiana	yes	KX381222
C1187	*Dermanura cinerea*	/	/	French Guiana	yes	KX381223
C1329	*Hsunycteris thomasi*	/	/	French Guiana	yes	KX381224
C1348	*Mimon bennettii*	/	/	French Guiana	no	KX381225
C1369	*Artibeus concolor*	/	/	French Guiana	yes	KX381226
C1370	*Carollia brevicauda*	/	/	French Guiana	no	KX381227
C1376	*Lophostoma silvicolum*	/	/	French Guiana	yes	KX381228
C1380	*Hsunycteris thomasi*	/	/	French Guiana	yes	KX381229
C1406	*Chiroderma villosum*	/	/	French Guiana	yes	KX381230
C1419	*Lonchorhina inusitata*	/	/	French Guiana	yes	KX381231
C1447	*Trinycteris nicefori*	/	/	French Guiana	yes	KX381232
C1448	*Dermanura gnoma*	/	/	French Guiana	yes	KX381233
C1450	*Artibeus lituratus*	/	/	French Guiana	yes	KX381234
C1467	*Trachops cirrhosus*	/	/	French Guiana	yes	KX381235
C1480	*Chiroderma villosum*	/	/	French Guiana	yes	KX381236
C1482	*Artibeus planirostris*	/	/	French Guiana	yes	KX381237
C1493	*Cormura brevirostris*	/	/	French Guiana	yes	KX381238
C1494	*Saccopteryx bilineata*	/	/	French Guiana	yes	KX381239
C1573	*Pteronotus rubiginosus*	/	/	French Guiana	yes	KX381240
C1576	*Choeroniscus minor*	/	/	French Guiana	yes	KX381241
C1590	*Micronycteris hirsuta*	/	/	French Guiana	yes	KX381242
C1614	*Dermanura gnoma*	/	/	French Guiana	yes	KX381243
C1641	*Carollia brevicauda*	/	/	French Guiana	yes	KX381244
C1644	*Lophostoma schulzi*	/	/	French Guiana	yes	KX381245
C1652	*Carollia brevicauda*	/	/	French Guiana	yes	KX381246
C1751	*Molossus barnesi*	/	/	French Guiana	yes	KX381247
C1752	*Molossus barnesi*	/	/	French Guiana	yes	KX381248
C1763	*Carollia perspicillata*	/	/	French Guiana	yes	KX381249
C1764	*Carollia perspicillata*	/	/	French Guiana	yes	KX381250
C1765	*Carollia perspicillata*	/	/	French Guiana	yes	KX381251
C1766	*Carollia perspicillata*	/	/	French Guiana	yes	KX381252
C1771	*Molossus molossus*	/	/	French Guiana	yes	KX381253
C1772	*Molossus molossus*	/	/	French Guiana	yes	KX381254
C1773	*Molossus molossus*	/	/	French Guiana	yes	KX381255
C1774	*Molossus molossus*	/	/	French Guiana	yes	KX381256
C1788	*Tonatia saurophila*	/	/	French Guiana	yes	KX381257
C1791	*Molossus barnesi*	/	/	French Guiana	yes	KX381258
C1792	*Furipterus horrens*	/	/	French Guiana	yes	KX381259
C1805	*Lophostoma silvicolum*	/	/	French Guiana	yes	KX381260
C1807	*Lophostoma silvicolum*	/	/	French Guiana	yes	KX381261
C1822	*Lophostoma silvicolum*	/	/	French Guiana	yes	KX381262
C1838	*Artibeus obscurus*	/	/	French Guiana	yes	KX381263
C1839	*Artibeus obscurus*	/	/	French Guiana	yes	KX381264
C1844	*Choeroniscus minor*	/	/	French Guiana	yes	KX381265
C1845	*Artibeus planirostris*	/	/	French Guiana	yes	KX381266
C1865	*Carollia brevicauda*	/	/	French Guiana	yes	KX381267

C1900	*Carollia perspicillata*	/	/	French Guiana	yes	KX381268
C1960	*Glyphonycteris sylvestris*	/	/	French Guiana	yes	KX381269
C1972	*Anoura geoffroyi*	/	/	French Guiana	yes	KX381270
C1983	*Pteronotus rubiginosus*	/	/	French Guiana	yes	KX381271
C1985	*Pteronotus rubiginosus*	/	/	French Guiana	yes	KX381272
C1989	*Trinycteris nicefori*	/	/	French Guiana	yes	KX381273
C2004	*Phylloderma stenops*	/	/	French Guiana	yes	KX381274
C2016	*Trinycteris nicefori*	/	/	French Guiana	yes	KX381275
C2018	*Phyllostomus latifolius*	/	/	French Guiana	yes	KX381276
C2041	*Lophostoma schulzi*	/	/	French Guiana	yes	KX381277
C2046	*Lophostoma schulzi*	/	/	French Guiana	yes	KX381278
C2079	*Micronycteris megalotis*	/	/	French Guiana	yes	KX381279
C2086	*Myotis riparius*	/	/	French Guiana	yes	KX381280
C2090	*Micronycteris megalotis*	/	/	French Guiana	yes	KX381281
C2094	*Phyllostomus latifolius*	/	/	French Guiana	yes	KX381282
C2106	*Chiroderma trinitatum*	/	/	French Guiana	yes	KX381283
C2111	*Lionycteris spurrelli*	/	/	French Guiana	yes	KX381284
C2143	*Pteronotus cf. sp.3*	/	/	French Guiana	yes	KX381285
C2144	*Pteronotus rubiginosus*	/	/	French Guiana	yes	KX381286
C2145	*Pteronotus cf. sp.3*	/	/	French Guiana	yes	KX381287
C2146	*Pteronotus rubiginosus*	/	/	French Guiana	yes	KX381288
C2147	*Pteronotus cf. sp.3*	/	/	French Guiana	yes	KX381289
C2148	*Pteronotus cf. sp.3*	/	/	French Guiana	yes	KX381290
C2159	*Diaemus youngii*	/	/	French Guiana	yes	KX381291
C2160	*Diaemus youngii*	/	/	French Guiana	yes	KX381292
C2161	*Diaemus youngii*	/	/	French Guiana	yes	KX381293
C2184	*Choeroniscus minor*	CM003	MNHN	French Guiana	yes	KX381294
C2186	*Dermanura gnoma*	ROM-113904	ROM	Suriname	yes	KX381295
C2187	*Dermanura gnoma*	ROM-113905	ROM	Suriname	no	KX381296
C2195	*Trachops cirrhosus*	V-2473	ISEM	French Guiana	yes	KX381300
C2234	*Desmodus rotundus*	/	/	French Guiana	yes	KX381308
C2235	*Desmodus rotundus*	/	/	French Guiana	yes	KX381309
C2236	*Desmodus rotundus*	/	/	French Guiana	yes	KX381310
C2237	*Desmodus rotundus*	/	/	French Guiana	yes	KX381311
C2250	*Molossus rufus*	/	/	French Guiana	yes	KX381312
C2254	*Myotis riparius*	/	/	French Guiana	yes	KX381313
C2255	*Mimon bennettii*	/	/	French Guiana	no	KX381314
C2276	*Cormura brevirostris*	/	/	French Guiana	yes	KX381315
C2277	*Mimon crenulatum*	/	/	French Guiana	yes	KX381316
C2278	*Chiroderma trinitatum*	/	/	French Guiana	yes	KX381317
C2280	*Chiroderma villosum*	/	/	French Guiana	yes	KX381318
C2296	*Platyrrhinus brachycephalus*	/	/	French Guiana	yes	KX381319
C2319	*Eumops auripendulus*	/	/	French Guiana	yes	KX381320
C2320	*Micronycteris megalotis*	/	/	French Guiana	yes	KX381321
C2321	*Vampyrum spectrum*	/	/	French Guiana	no	KX381322
C2346	*Thyroptera discifera*	/	/	French Guiana	no	KX381323
C2354	*Saccopteryx leptura*	/	/	French Guiana	yes	KX381324
C2356	*Platyrrhinus fusciventris*	/	/	French Guiana	yes	KX381325
C2359	*Platyrrhinus fusciventris*	/	/	French Guiana	yes	KX381326
C2368	*Phyllostomus latifolius*	/	/	French Guiana	no	KX381327
C2374	*Saccopteryx bilineata*	/	/	French Guiana	yes	KX381328
C2382	*Saccopteryx bilineata*	/	/	French Guiana	yes	KX381329
C2389	*Ametrida centurio*	/	/	French Guiana	yes	KX381330
C2398	*Noctilio leporinus*	/	/	French Guiana	yes	KX381331
C2399	*Cynomops planirostris*	/	/	French Guiana	yes	KX381332
C2401	*Cynomops paranus*	/	/	French Guiana	yes	KX381333
C2402	*Eptesicus furinalis*	/	/	French Guiana	yes	KX381334
C2403	*Cynomops paranus*	/	/	French Guiana	yes	KX381335
C2404	*Cynomops planirostris*	/	/	French Guiana	yes	KX381336
C2405	*Cynomops planirostris*	/	/	French Guiana	no	KX381337
C2406	*Cynomops planirostris*	/	/	French Guiana	no	KX381338
C2407	*Cynomops paranus*	/	/	French Guiana	yes	KX381339
C2409	*Cynomops planirostris*	/	/	French Guiana	yes	KX381340
C2413	*Cynomops abrasus*	/	/	French Guiana	yes	KX381341
C2414	*Cynomops abrasus*	/	/	French Guiana	yes	KX381342
C2415	*Cynomops abrasus*	/	/	French Guiana	yes	KX381343
C2416	*Cynomops abrasus*	/	/	French Guiana	yes	KX381344
C2418	*Molossus rufus*	/	/	French Guiana	yes	KX381345
C2422	*Eptesicus furinalis*	/	/	French Guiana	yes	KX381346
C2424	*Noctilio albiventris*	/	/	French Guiana	yes	KX381347
C2425	*Noctilio leporinus*	/	/	French Guiana	yes	KX381348
C2428	*Noctilio albiventris*	/	/	French Guiana	no	KX381349
C2435	*Noctilio leporinus*	/	/	French Guiana	no	KX381350
C2437	*Noctilio leporinus*	/	/	French Guiana	yes	KX381351
C2441	*Eptesicus furinalis*	/	/	French Guiana	yes	KX381352
C2443	*Phyllostomus hastatus*	/	/	French Guiana	no	KX381353

C2446	*Rhinophylla pumilio*	/	/	French Guiana	yes	KX381354
C2452	*Rhinophylla pumilio*	/	/	French Guiana	yes	KX381355
C2454	*Rhinophylla pumilio*	/	/	French Guiana	yes	KX381356
C2461	*Saccopteryx leptura*	/	/	French Guiana	yes	KX381357
C2466	*Artibeus obscurus*	/	/	French Guiana	yes	KX381358
C2480	*Chrotopterus auritus*	/	/	French Guiana	yes	KX381360
C2482	*Natalus tumidirostris*	/	/	French Guiana	yes	KX381361
C2487	*Saccopteryx leptura*	/	/	French Guiana	yes	KX381362
C2489	*Micronycteris brosseti*	/	/	French Guiana	yes	KX381363
C2497	*Molossus rufus*	/	/	French Guiana	yes	KX381364
C2504	*Lophostoma silvicolum*	/	/	French Guiana	yes	KX381366
C2507	*Molossus rufus*	/	/	French Guiana	yes	KX381367
C2512	*Phyllostomus hastatus*	/	/	French Guiana	yes	KX381369
C2594	*Micronycteris minuta*	/	/	French Guiana	yes	KX381371
C2697	*Uroderma cf. magnirostrum*	/	/	French Guiana	yes	KX381372
C2718	*Vampyrodes caraccioli*	/	/	French Guiana	yes	KX381373
C2720	*Lonchorhina inusitata*	/	/	French Guiana	yes	KX381374
C2723	*Mimon bennettii*	/	/	French Guiana	no	KX381375
C2725	*Glyphonycteris daviesi*	/	/	French Guiana	yes	KX381376
C2727	*Phyllostomus elongatus*	/	/	French Guiana	yes	KX381377
C2728	*Phyllostomus elongatus*	/	/	French Guiana	yes	KX381378
C2732	*Hsunycteris thomasi*	/	/	French Guiana	yes	KX381379
C2739	*Phyllostomus elongatus*	/	/	French Guiana	yes	KX381380
C2765	*Mesophylla macconnelli*	/	/	French Guiana	yes	KX381381
C2767	*Mesophylla macconnelli*	/	/	French Guiana	yes	KX381382
C2772	*Phyllostomus latifolius*	/	/	French Guiana	yes	KX381383
C2799	*Micronycteris hirsuta*	/	/	French Guiana	yes	KX381384
C2807	*Saccopteryx leptura*	/	/	French Guiana	yes	KX381385
C2813	*Phyllostomus hastatus*	/	/	French Guiana	no	KX381386
C2825	*Rhinophylla pumilio*	/	/	French Guiana	yes	KX381387
C2827	*Rhinophylla pumilio*	/	/	French Guiana	yes	KX381388
C2828	*Mesophylla macconnelli*	/	/	French Guiana	yes	KX381389
C2829	*Mesophylla macconnelli*	/	/	French Guiana	yes	KX381390
C2830	*Mesophylla macconnelli*	/	/	French Guiana	yes	KX381391
C2845	*Micronycteris hirsuta*	/	/	French Guiana	yes	KX381392
C2846	*Ametrida centurio*	/	/	French Guiana	yes	KX381393
C2848	*Phylloderma stenops*	/	/	French Guiana	yes	KX381394
C2856	*Mimon crenulatum*	/	/	French Guiana	yes	KX381395
C2864	*Artibeus obscurus*	/	/	French Guiana	yes	KX381396
C547	*Artibeus concolor*	/	/	French Guiana	yes	KX381397
C614	*Artibeus planirostris*	/	/	French Guiana	yes	KX381398
C840	*Uroderma bilobatum*	/	/	French Guiana	no	KX381399
C844	*Eptesicus furinalis*	/	/	French Guiana	yes	KX381400
C92	*Phyllostomus discolor*	/	/	French Guiana	yes	KX381401
C960	*Sturnira sp.3*	/	/	French Guiana	yes	KX381402
C968	*Sturnira sp.3*	/	/	French Guiana	yes	KX381403
FBC11	*Chrotopterus auritus*	AMNH-109773	AMNH	Peru	no	KX381404
FBC15	*Diaemus youngii*	NK-13129	MSB	Bolivia	yes	KX381405
FBC18	*Glyphonycteris sylvestris*	MUSM-15202	AMNH	Peru	yes	KX381406
FBC2	*Anoura caudifer*	ROM-113962	ROM	Suriname	yes	KX381407
FBC25	*Macrophyllum macrophyllum*	NK-13979	MSB	Bolivia	yes	KX381408
FBC26	*Mesophylla macconnelli*	AMNH-272725	AMNH	Peru	yes	KX381409
FBC46	*Vampyriscus bidens*	NK-13660	MSB	Bolivia	yes	KX381410
M1015	*Eira barbara*	/	/	French Guiana	yes	KX381411
M1017	*Tapirus terrestris*	/	/	French Guiana	yes	KX381412
M1018	*Tapirus terrestris*	/	/	French Guiana	yes	KX381413
M1020	*Tapirus terrestris*	/	/	French Guiana	yes	KX381414
M1021	*Tapirus terrestris*	/	/	French Guiana	yes	KX381415
M1085	*Alouatta macconnelli*	/	/	French Guiana	yes	KX381416
M1086	*Pithecia pithecia*	/	/	French Guiana	yes	KX381417
M1087	*Ateles paniscus*	/	/	French Guiana	yes	KX381418
M1093	*Oecomys rutilus*	/	/	French Guiana	no	KX381419
M1097	*Mazama nemorivaga*	/	/	French Guiana	yes	KX381420
M1106	*Cebus apella*	/	/	French Guiana	yes	KX381421
M1112	*Neacomys paracou*	/	/	French Guiana	yes	KX381422
M1114	*Agouti paca*	/	/	French Guiana	yes	KX381423
M1121	*Agouti paca*	/	/	French Guiana	yes	KX381424
M1125	*Agouti paca*	/	/	French Guiana	yes	KX381425
M1126	*Euryoryzomys macconnelli*	/	/	French Guiana	yes	KX381426
M1129	*Dasyprocta leporina*	/	/	French Guiana	yes	KX381427
M1131	*Leopardus pardalis*	/	/	French Guiana	yes	KX381428
M1136	*Neacomys paracou*	/	/	French Guiana	yes	KX381429
M1137	*Agouti paca*	/	/	French Guiana	yes	KX381430
M1141	*Dasyprocta leporina*	/	/	French Guiana	yes	KX381431
M1154	*Eira barbara*	/	/	French Guiana	yes	KX381432
M1156	*Marmosa murina*	/	/	French Guiana	yes	KX381433

M1160	*Proechimys cuvieri*	/	/	French Guiana	yes	KX381434
M1161	*Marmosa murina*	/	/	French Guiana	yes	KX381435
M1162	*Proechimys cuvieri*	/	/	French Guiana	yes	KX381436
M119	*Pithecia pithecia*	/	/	French Guiana	yes	KX381439
M1211	*Didelphis marsupialis*	/	/	French Guiana	yes	KX381440
M1268	*Rhipidomys nitela*	/	/	French Guiana	yes	KX381441
M1309	*Proechimys cuvieri*	/	/	French Guiana	yes	KX381442
M1358	*Rattus rattus*	/	/	French Guiana	yes	KX381445
M1359	*Rattus rattus*	/	/	French Guiana	yes	KX381446
M1362	*Coendou prehensilis*	/	/	French Guiana	yes	KX381447
M1399	*Ateles paniscus*	/	/	French Guiana	yes	KX381449
M1414	*Pteronura brasiliensis*	/	/	Peru	no	KX381451
M1415	*Pteronura brasiliensis*	/	/	Brazil	yes	KX381452
M1416	*Pteronura brasiliensis*	/	/	Peru	yes	KX381453
M1420	*Pteronura brasiliensis*	/	/	Colombia	yes	KX381454
M1421	*Pteronura brasiliensis*	/	/	Colombia	yes	KX381455
M143	*Oecomys bicolor*	/	/	Peru	yes	KX381456
M15	*Choloepus didactylus*	/	/	French Guiana	yes	KX381461
M1553	*Saguinus midas*	/	/	French Guiana	yes	KX381463
M1572	*Puma yagouaroundi*	/	/	French Guiana	yes	KX381465
M1574	*Coendou prehensilis*	/	/	French Guiana	yes	KX381466
M1576	*Ateles paniscus*	/	/	French Guiana	yes	KX381467
M1577	*Ateles paniscus*	/	/	French Guiana	yes	KX381468
M158	*Marmosops pinheiroi*	/	/	French Guiana	yes	KX381469
M164	*Rhipidomys nitela*	/	/	French Guiana	yes	KX381470
M1665	*Hylaeamys megacephalus*	/	/	French Guiana	yes	KX381472
M1674	*Neacomys paracou*	/	/	French Guiana	yes	KX381473
M1680	*Coendou prehensilis*	/	/	French Guiana	yes	KX381475
M1681	*Cabassous unicinctus*	/	/	French Guiana	yes	KX381476
M1686	*Cyclopes didactylus*	/	/	French Guiana	yes	KX381477
M1687	*Cabassous unicinctus*	/	/	French Guiana	yes	KX381478
M1694	*Puma yagouaroundi*	/	/	French Guiana	yes	KX381479
M1699	*Caluromys philander*	/	/	French Guiana	yes	KX381480
M170	*Caluromys philander*	/	/	French Guiana	yes	KX381481
M1706	*Alouatta macconnelli*	/	/	French Guiana	yes	KX381482
M1721	*Saguinus midas*	/	/	French Guiana	yes	KX381483
M1722	*Saguinus midas*	/	/	French Guiana	yes	KX381484
M1737	*Pithecia pithecia*	/	/	French Guiana	yes	KX381485
M1749	*Potos flavus*	/	/	French Guiana	yes	KX381487
M1753	*Dasypus kappleri*	/	/	French Guiana	no	KX381488
M1755	*Speothos venaticus*	/	/	French Guiana	no	KX381489
M1756	*Speothos venaticus*	/	/	Brazil	yes	KX381490
M1757	*Chiropotes chiropotes*	/	/	Brazil	yes	KX381491
M176	*Caluromys philander*	/	/	French Guiana	yes	KX381492
M1773	*Eira barbara*	/	/	French Guiana	yes	KX381493
M1780	*Potos flavus*	/	/	French Guiana	yes	KX381494
M1786	*Potos flavus*	/	/	French Guiana	yes	KX381495
M185	*Coendou prehensilis*	/	/	French Guiana	yes	KX381496
M1859	*Dasypus kappleri*	/	/	French Guiana	yes	KX381497
M1860	*Coendou melanurus*	/	/	French Guiana	no	KX381498
M1861	*Bradypus tridactylus*	/	/	French Guiana	yes	KX381499
M1862	*Coendou melanurus*	/	/	French Guiana	yes	KX381500
M1863	*Bradypus tridactylus*	/	/	French Guiana	yes	KX381501
M1864	*Dasypus kappleri*	/	/	French Guiana	yes	KX381502
M1866	*Puma yagouaroundi*	/	/	French Guiana	yes	KX381503
M1867	*Puma yagouaroundi*	/	/	French Guiana	yes	KX381504
M1868	*Leopardus wiedii*	/	/	French Guiana	yes	KX381505
M1869	*Leopardus wiedii*	/	/	French Guiana	yes	KX381506
M187	*Eira barbara*	/	/	French Guiana	yes	KX381507
M1873	*Leopardus pardalis*	/	/	French Guiana	yes	KX381508
M1875	*Panthera onca*	/	/	French Guiana	yes	KX381509
M1879	*Leopardus pardalis*	/	/	French Guiana	yes	KX381510
M1881	*Coendou melanurus*	/	/	French Guiana	no	KX381511
M190	*Rattus norvegicus*	/	/	French Guiana	no	KX381512
M191	*Procyon cancrivorus*	/	/	French Guiana	yes	KX381513
M192	*Procyon cancrivorus*	/	/	French Guiana	yes	KX381514
M194	*Hydrochoerus hydrochaeris*	/	/	French Guiana	yes	KX381515
M2	*Choloepus didactylus*	/	/	French Guiana	yes	KX381516
M21	*Tamandua tetradactyla*	/	/	French Guiana	no	KX381517
M227	*Mazama americana*	/	/	French Guiana	yes	KX381518
M228	*Mazama nemorivaga*	/	/	French Guiana	yes	KX381519
M256	*Mazama americana*	/	/	French Guiana	yes	KX381520
M26	*Cabassous unicinctus*	/	/	French Guiana	yes	KX381521
M264	*Marmosops parvidens*	/	/	French Guiana	yes	KX381522
M266	*Makalata didelphoides*	/	/	French Guiana	yes	KX381523
M268	*Makalata didelphoides*	/	/	French Guiana	yes	KX381524

M27	*Myoprocta acouchy*	/	/	French Guiana	yes	KX381525
M275	*Makalata didelphoides*	/	/	French Guiana	yes	KX381526
M278	*Dasypus novemcinctus*	/	/	French Guiana	yes	KX381527
M288	*Dasypus novemcinctus*	/	/	French Guiana	yes	KX381528
M290	*Procyon cancrivorus*	/	/	French Guiana	yes	KX381529
M295	*Dasypus novemcinctus*	/	/	French Guiana	yes	KX381530
M33	*Panthera onca*	/	/	French Guiana	yes	KX381533
M330	*Marmosa demerarae*	/	/	French Guiana	yes	KX381534
M337	*Marmosops parvidens*	/	/	French Guiana	yes	KX381537
M348	*Trichechus manatus*	/	/	French Guiana	yes	KX381539
M350	*Didelphis imperfecta*	/	/	French Guiana	yes	KX381540
M351	*Oecomys sp-1*	/	/	French Guiana	yes	KX381541
M363	*Marmosops parvidens*	/	/	French Guiana	yes	KX381542
M365	*Makalata didelphoides*	/	/	French Guiana	yes	KX381543
M367	*Hylaeamys yunganus*	/	/	French Guiana	yes	KX381544
M416	*Monodelphis touan*	/	/	French Guiana	yes	KX381551
M42	*Galictis vittata*	/	/	French Guiana	yes	KX381552
M423	*Trichechus manatus*	/	/	French Guiana	yes	KX381553
M43	*Galictis vittata*	/	/	French Guiana	yes	KX381554
M44	*Galictis vittata*	/	/	French Guiana	yes	KX381555
M495	*Marmosops pinheiroi*	/	/	French Guiana	yes	KX381556
M50	*Lontra longicaudis*	/	/	French Guiana	yes	KX381558
M503	*Didelphis marsupialis*	/	/	French Guiana	yes	KX381559
M507	*Oecomys sp-1*	/	/	French Guiana	yes	KX381560
M51	*Lontra longicaudis*	/	/	French Guiana	yes	KX381561
M524	*Trichechus manatus*	/	/	French Guiana	yes	KX381562
M525	*Dasypus novemcinctus*	/	/	French Guiana	yes	KX381563
M53	*Lontra longicaudis*	/	/	French Guiana	yes	KX381564
M555	*Oecomys auyantepui*	/	/	French Guiana	yes	KX381565
M572	*Hylaeamys megacephalus*	/	/	French Guiana	yes	KX381567
M573	*Marmosops pinheiroi*	/	/	French Guiana	yes	KX381568
M575	*Hylaeamys megacephalus*	/	/	French Guiana	yes	KX381569
M576	*Hylaeamys yunganus*	/	/	French Guiana	yes	KX381570
M580	*Didelphis imperfecta*	/	/	French Guiana	yes	KX381571
M581	*Didelphis imperfecta*	/	/	French Guiana	yes	KX381572
M604	*Choloepus didactylus*	/	/	French Guiana	yes	KX381573
M606	*Oecomys auyantepui*	/	/	French Guiana	yes	KX381574
M620	*Galictis vittata*	/	/	French Guiana	yes	KX381575
M626	*Procyon cancrivorus*	/	/	French Guiana	yes	KX381576
M65	*Myoprocta acouchy*	/	/	French Guiana	yes	KX381577
M70	*Odocoileus cariacou*	/	/	French Guiana	no	KX381578
M71	*Odocoileus cariacou*	/	/	French Guiana	no	KX381579
M719	*Marmosa demerarae*	/	/	French Guiana	yes	KX381580
M730	*Didelphis marsupialis*	/	/	French Guiana	yes	KX381581
M75	*Myoprocta acouchy*	/	/	French Guiana	yes	KX381582
M752	*Oecomys sp-1*	/	/	French Guiana	yes	KX381583
M79	*Hydrochoerus hydrochaeris*	/	/	French Guiana	yes	KX381584
M791	*Marmosa demerarae*	/	/	French Guiana	yes	KX381585
M824	*Coendou melanurus*	/	/	French Guiana	yes	KX381586
M844	*Priodontes maximus*	/	/	French Guiana	yes	KX381588
M867	*Cebus apella*	/	/	French Guiana	yes	KX381589
M872	*Myoprocta acouchy*	/	/	French Guiana	yes	KX381590
M873	*Saimiri sciureus*	/	/	French Guiana	yes	KX381591
M879	*Mazama nemorivaga*	/	/	French Guiana	yes	KX381592
M908	*Oligoryzomys fulvescens*	/	/	French Guiana	yes	KX381597
M917	*Leopardus wiedii*	/	/	French Guiana	yes	KX381602
M918	*Metachirus nudicaudatus*	/	/	French Guiana	yes	KX381603
M919	*Saguinus midas*	/	/	French Guiana	yes	KX381604
M949	*Alouatta macconnelli*	/	/	French Guiana	yes	KX381613
M954	*Cebus apella*	/	/	French Guiana	yes	KX381614
M955	*Alouatta macconnelli*	/	/	French Guiana	yes	KX381615
M98	*Tamandua tetradactyla*	/	/	French Guiana	yes	KX381620
R229	*Artibeus planirostris*	/	/	French Guiana	yes	KX381622
R246	*Artibeus planirostris*	/	/	French Guiana	yes	KX381623
R259	*Artibeus planirostris*	/	/	French Guiana	yes	KX381624
R98	*Phyllostomus discolor*	/	/	French Guiana	yes	KX381625
SC127	*Anoura geoffroyi*	/	/	French Guiana	yes	KX381626
SC128	*Anoura geoffroyi*	/	/	French Guiana	yes	KX381627
T-1069	*Panthera onca*	/	Zoo de la Palmyre	France	no	KX381628
T-1254	*Puma concolor*	Z80180	Zoo de Vincennes	France	no	KX381629
T-1399	*Euryoryzomys macconnelli*	MNHN-M&O-1994-126	MNHN	French Guiana	yes	KX381630
T-1479	*Agouti paca*	MNHN-M&O-1997-646	MNHN	French Guiana	no	KX381631
T-1487	*Coendou melanurus*	MNHN-M&O-1997-640	MNHN	French Guiana	no	KX381632
T-1557	*Bradypus tridactylus*	/	/	French Guiana	no	KX381633
T-1560	*Nasua nasua*	MNHN-M&O-1999-1072	MNHN	French Guiana	no	KX381634
T-1562	*Mazama americana*	MNHN-M&O-1999-1076	MNHN	French Guiana	yes	KX381635

T-1563	*Mazama nemorivaga*	MNHN-M&O-1999-1077	MNHN	French Guiana	no	KX381636
T-1607	*Dasypus novemcinctus*	MNHN-M&O-1999-1069	MNHN	French Guiana	yes	KX381637
T-1618	*Hydrochoerus hydrochaeris*	/	/	French Guiana	yes	KX381638
T-1625	*Alouatta macconnelli*	MNHN-M&O-1998-235	MNHN	French Guiana	yes	KX381639
T-1626	*Coendou prehensilis*	MNHN-M&O-1997-643	MNHN	French Guiana	yes	KX381640
T-1631	*Cyclopes didactylus*	MNHN-M&O-1998-234	MNHN	French Guiana	yes	KX381641
T-1636	*Eira barbara*	MNHN-M&O-1998-2256	MNHN	French Guiana	yes	KX381642
T-1641	*Cabassous unicinctus*	MNHN-M&O-1999-1068	MNHN	French Guiana	no	KX381643
T-1722	*Choloepus didactylus*	MNHN-M&O-1998-1819	MNHN	French Guiana	yes	KX381644
T-1764	*Trichechus manatus*	MNHN-M&O-1998-1865	MNHN	French Guiana	yes	KX381645
T-1820	*Philander opossum*	MNHN-M&O-2000-215	MNHN	French Guiana	yes	KX381646
T-1826	*Sotalia guianensis*	/	/	French Guiana	yes	KX381647
T-1857	*Proechimys cuvieri*	MNHN-M&O-1999-1094	MNHN	French Guiana	no	KX381648
T-1862	*Saimiri sciureus*	MNHN-M&O-2000-1052	MNHN	French Guiana	yes	KX381649
T-1863	*Dasypus novemcinctus*	/	ECOBIOS	French Guiana	yes	KX381650
T-1910	*Vampyrodes caraccioli*	PCD-1053	MNHN (Brunoy)	French Guiana	yes	KX381651
T-1972	*Tayassu pecari*	/	/	French Guiana	no	KX381652
T-1978	*Pecari tajacu*	/	/	French Guiana	yes	KX381653
T-1995	*Lontra longicaudis*	MNHN-M&O-2001-1970	MNHN	French Guiana	yes	KX381654
T-2252	*Pecari tajacu*	/	/	Suriname	yes	KX381655
T-2291	*Cabassous unicinctus*	/	/	French Guiana	yes	KX381656
T-2378	*Tayassu pecari*	/	/	French Guiana	yes	KX381657
T-2387	*Cebus apella*	/	/	French Guiana	no	KX381658
T-2404	*Potos flavus*	/	/	French Guiana	yes	KX381659
T-2447	*Pecari tajacu*	/	/	French Guiana	yes	KX381660
T-2480	*Myoprocta acouchy*	V-1187	ISEM	French Guiana	yes	KX381661
T-2511	*Neacomys dubosti*	V-1128	ISEM	French Guiana	no	KX381662
T-2521	*Oecomys rex*	MNHN-M&O-2003-38	MNHN	French Guiana	yes	KX381663
T-2524	*Neacomys dubosti*	V-1141	ISEM	French Guiana	yes	KX381664
T-2582	*Dermanura cinerea*	PCD-1118	MNHN (Brunoy)	French Guiana	yes	KX381665
T-2862	*Myrmecophaga tridactyla*	/	/	French Guiana	yes	KX381666
T-2930	*Tayassu pecari*	/	/	French Guiana	yes	KX381667
T-2946	*Tayassu pecari*	/	/	French Guiana	yes	KX381668
T-2971	*Tayassu pecari*	/	/	French Guiana	yes	KX381669
T-3295	*Mesomys hispidus*	MHNG-1885.021	MHNG	French Guiana	yes	KX381670
T-3296	*Oecomys bicolor*	MHNG-1889.089	MHNG	French Guiana	yes	KX381671
T-3365	*Dasypus kappleri*	/	/	French Guiana	yes	KX381672
T-3393	*Vampyrum spectrum*	L-1896	EBRG	Venezuela	yes	KX381673
T-3418	*Desmodus rotundus*	L-1874	EBRG	Venezuela	yes	KX381674
T-3430	*Ametrida centurio*	L-1897	EBRG	Venezuela	no	KX381675
T-3477	*Micronycteris hirsuta*	1052-PCD	MNHN (Brunoy)	French Guiana	yes	KX381676
T-3667	*Vampyrodes caraccioli*	F-40993	ROM	Suriname	no	KX381677
T-3711	*Glyphonycteris daviesi*	F-41125-(ROM)	ROM	Suriname	Yes	KX381678
T-3715	*Phyllostomus latifolius*	F-41138-(ROM)	ROM	Suriname	Yes	KX381679
T-3722	*Choeroniscus minor*	F-41181-(ROM)	ROM	Suriname	no	KX381680
T-3726	*Glyphonycteris daviesi*	F-41189	ROM	Suriname	no	KX381681
T-3727	*Lophostoma schulzi*	F-41190	ROM	Suriname	no	KX381682
T-3748	*Chiroderma trinitatum*	F-41245	ROM	Suriname	yes	KX381683
T-3759	*Chiroderma villosum*	F-41287	ROM	Suriname	yes	KX381684
T-3765	*Vampyriscus brocki*	F-41310	ROM	Suriname	yes	KX381685
T-3843	*Myotis riparius*	PCD-1244	MNHN (Brunoy)	French Guiana	yes	KX381686
T-3849	*Dermanura gnoma*	PCD-1246	MNHN (Brunoy)	French Guiana	no	KX381687
T-3852	*Peropteryx kappleri*	PCD-1245	MNHN (Brunoy)	French Guiana	no	KX381688
T-4107	*Pecari tajacu*	/	/	French Guiana	no	KX381689
T-4109	*Pecari tajacu*	/	/	French Guiana	yes	KX381690
T-4115	*Tayassu pecari*	/	/	French Guiana	yes	KX381691
T-4253	*Pithecia pithecia*	/	/	French Guiana	yes	KX381692
T-4333	*Tayassu pecari*	/	/	French Guiana	no	KX381693
T-4342	*Leopardus wiedii*	/	/	French Guiana	no	KX381694
T-4366	*Cynomops abrasus*	MHNG-1880.044	MHNG	French Guiana	yes	KX381695
T-4369	*Lichonycteris obscura*	MHNG-1880.041	MHNG	French Guiana	yes	KX381696
T-4377	*Isothrix sinnamariensis*	V-1708	ISEM	French Guiana	yes	KX381697
T-4385	*Hyladelphys kalinowskii*	V-1791	ISEM	French Guiana	yes	KX381698
T-4391	*Myotis riparius-PS1*	MHNG-1885.024	MHNG	French Guiana	yes	KX381699
T-4393	*Macrophyllum macrophyllum*	MHNG-1885.031	MHNG	French Guiana	yes	KX381700
T-4411	*Sturnira sp.3*	MHNG-1885.027	MHNG	French Guiana	yes	KX381701
T-4427	*Galictis vittata*	/	/	French Guiana	no	KX381702
T-4450	*Glossophaga soricina*	MNHN-M&O-2004-350	MNHN	French Guiana	no	KX381703
T-4455	*Trinycteris nicefori*	MNHN-M&O-2004-378	MNHN	French Guiana	yes	KX381704
T-4458	*Phyllostomus discolor*	MNHN-M&O-2004-359	MNHN	French Guiana	yes	KX381705
T-4481	*Phyllostomus elongatus*	MNHN-M&O-2004-360	MNHN	French Guiana	yes	KX381706
T-4488	*Tonatia saurophila*	MNHN-M&O-2004-376	MNHN	French Guiana	yes	KX381707
T-4493	*Saccopteryx bilineata*	MNHN-M&O-2004-366	MNHN	French Guiana	yes	KX381708
T-4496	*Artibeus planirostris*	MNHN-M&O-2004-336	MNHN	French Guiana	yes	KX381709
T-4497	*Lophostoma silvicolum*	MNHN-M&O-2004-352	MNHN	French Guiana	no	KX381710

T-4500	*Phylloderma stenops*	MNHN-M&O-2004-358	MNHN	French Guiana	yes	KX381711
T-4573	*Noctilio leporinus*	MHNG-1939.065	MHNG	French Guiana	yes	KX381712
T-4584	*Eumops auripendulus*	MHNG-1939.070	MHNG	French Guiana	yes	KX381713
T-4591	*Noctilio leporinus*	MHNG-1939.075	MHNG	French Guiana	no	KX381714
T-4614	*Gracilinanus emiliae*	V-1932	ISEM	French Guiana	yes	KX381715
T-4663	*Tonatia saurophila*	V-1985	ISEM	French Guiana	yes	KX381716
T-4670	*Rhinophylla pumilio*	V-1992	ISEM	French Guiana	yes	KX381717
T-5013	*Bradypus tridactylus*	/	/	French Guiana	yes	KX381718
T-5022	*Makalata didelphoides*	MHNG-1960.002	MHNG	French Guiana	yes	KX381719
T-5065	*Platyrrhinus incarum*	V-2180	ISEM	French Guiana	yes	KX381720
T-5151	*Procyon cancrivorus*	/	/	French Guiana	no	KX381721
T-5180	*Lophostoma brasiliense*	V-2287	ISEM	French Guiana	yes	KX381722
T-5191	*Anoura geoffroyi*	V-2298	ISEM	French Guiana	no	KX381723
T-5204	*Uroderma bilobatum*	MNHN-M&O-2008-1493	MNHN	French Guiana	no	KX381724
T-5215	*Pteronotus rubiginosus*	MHNG-1980.092 (V-2322)	MHNG	French Guiana	no	KX381725
T-5216	*Carollia brevicauda*	V-2323	ISEM	French Guiana	yes	KX381726
T-5230	*Sciurus aestuans*	MHNG-1963.031	MHNG	French Guiana	yes	KX381727
T-5280	*Didelphis imperfecta*	MNHN-M&O-2008-1057	MNHN	French Guiana	yes	KX381728
T-5282	*Oecomys rutilus*	MHNG-1969.030	MHNG	French Guiana	Yes	KX381729
T-5299	*Sturnira tildae*	V-2412	ISEM	French Guiana	yes	KX381730
T-5361	*Monodelphis touan*	MHNG-1969.082	MHNG	French Guiana	no	KX381731
T-5375	*Trachops cirrhosus*	V-2473	ISEM	French Guiana	no	KX381732
T-5381	*Caluromys philander*	MHNG-1969.098	MHNG	French Guiana	no	KX381733
T-5415	*Platyrrhinus brachycephalus*	MHNG-1979.021	MHNG	French Guiana	yes	KX381734
T-5430	*Artibeus concolor*	MNHN-M&O-2008-1511	MNHN	French Guiana	yes	KX381735
T-5488	*Chironectes minimus*	/	/	French Guiana	yes	KX381736
T-5541	*Eptesicus chiriquinus*	V-2622	ISEM	French Guiana	yes	KX381737
T-5638	*Echimys chrysurus*	/	/	French Guiana	no	KX381738
T-5649	*Rattus norvegicus*	MHNG-1975.025	MHNG	French Guiana	no	KX381739
T-5685	*Micronycteris minuta*	V-2803	ISEM	French Guiana	no	KX381740
T-5691	*Rattus rattus*	MHNG-1975.069	MHNG	French Guiana	no	KX381741
T-5709	*Dasyprocta leporina*	MHNG-1975.087	MHNG	French Guiana	no	KX381742
T-5710	*Rhipidomys nitela*	MHNG-1975.088	MHNG	French Guiana	no	KX381743
T-5718	*Pecari tajacu*	/	/	French Guiana	no	KX381744
T-5723	*Ateles paniscus*	/	/	French Guiana	no	KX381745
T-5757	*Leopardus pardalis*	/	/	French Guiana	no	KX381746
T-5760	*Neusticomys oyapocki*	MHNG-1978.042	MHNG	French Guiana	yes	KX381747
T-5767	*Nectomys rattus*	MHNG-1978.049	MHNG	French Guiana	no	KX381748
T-6021	*Marmosops parvidens*	MHNG-1974.032	MHNG	French Guiana	yes	KX381749
T-6054	*Tamandua tetradactyla*	/	/	French Guiana	no	KX381750
T-6119	*Marmosa demerarae*	MHNG-1979.039	MHNG	French Guiana	yes	KX381751
T-6171	*Mus musculus*	MHNG-1979.062	MHNG	French Guiana	yes	KX381752
T-6177	*Hylaeamys megacephalus*	MHNG-1979.068	MHNG	French Guiana	yes	KX381753
T-6181	*Neacomys paracou*	MHNG-1979.072	MHNG	French Guiana	yes	KX381754
T-6203	*Proechimys guyannensis*	MHNG-1979.091	MHNG	French Guiana	no	KX381755
T-6210	*Odocoileus cariacou*	/	/	French Guiana	yes	KX381756
T-6211	*Tapirus terrestris*	/	/	French Guiana	yes	KX381757
T-6245	*Molossus rufus*	MHNG-1990.009	MHNG	French Guiana	yes	KX381758
T-6248	*Mimon crenulatum*	MHNG-1990.012	MHNG	French Guiana	yes	KX381759
T-6520	*Thyroptera tricolor*	MHNG-1983.030	MHNG	French Guiana	yes	KX381760
T-6534	*Hsunycteris thomasi*	MHNG-1983.045	MHNG	French Guiana	yes	KX381761
T-6537	*Saccopteryx bilineata*	MHNG-1983.048	MHNG	French Guiana	yes	KX381762
T-6540	*Phyllostomus hastatus*	MHNG-1983.051	MHNG	French Guiana	yes	KX381763
T-6542	*Saccopteryx leptura*	MHNG-1983.053	MHNG	French Guiana	yes	KX381764
T-6547	*Pteronotus rubiginosus*	MHNG-1983.058	MHNG	French Guiana	no	KX381765
T-6549	*Mimon bennettii*	MHNG-1983.060	MHNG	French Guiana	no	KX381766
T-6557	*Cormura brevirostris*	MHNG-1983.068	MHNG	French Guiana	yes	KX381767
T-6562	*Choeroniscus minor*	MHNG-1983.073	MHNG	French Guiana	yes	KX381768
T-6583	*Rhynchonycteris naso*	MHNG-1983.095	MHNG	French Guiana	yes	KX381769
T-6586	*Anoura caudifer*	MHNG-1983.098	MHNG	French Guiana	no	KX381770
T-6588	*Ametrida centurio*	MHNG-1983.100	MHNG	French Guiana	yes	KX381771
T-6590	*Marmosa murina*	MHNG-1984.002	MHNG	French Guiana	yes	KX381772
T-6604	*Molossus molossus*	MHNG-1984.016	MHNG	French Guiana	no	KX381773
T-6605	*Glossophaga soricina*	MHNG-1984.017	MHNG	French Guiana	no	KX381774
T-6606	*Didelphis marsupialis*	MHNG-1984.018	MHNG	French Guiana	no	KX381775
T-6609	*Saguinus midas*	/	/	French Guiana	no	KX381776
T-6747	*Cebus apella*	/	/	French Guiana	yes	KX381777
T-6811	*Natalus tumidirostris*	MHNG-1990.029	MHNG	French Guiana	no	KX381778
T-6835	*Trachops cirrhosus*	MHNG-1991.027	MHNG	French Guiana	yes	KX381779
T-6837	*Hylaeamys yunganus*	MHNG-1991.004	MHNG	French Guiana	yes	KX381780
T-6872	*Zygodontomys brevicauda*	MHNG-1991.032	MHNG	French Guiana	yes	KX381781
T-6876	*Holochilus sciureus*	MHNG-1991.003	MHNG	French Guiana	yes	KX381782
TTA001	*Pecari tajacu*	/	/	French Guiana	yes	KX381783
V-3050	*Molossus barnesi*	MHNG-1983.020	MHNG	French Guiana	Yes	KX381784
C2193	*Vampyriscus brocki*	MNHN-M&O-2004-385	MNHN	French Guiana	yes	KX381299
C2197	*Dermanura gnoma*	MNHN-M&O-2008-1508	MNHN	French Guiana	yes	KX381302

C2198	*Dermanura gnoma*	MNHN-M&O-2008-1512	MNHN	French Guiana	yes	KX381303
C2202	*Glossophaga soricina*	MHNG-1991.002	MHNG	French Guiana	yes	KX381304
C2212	*Micronycteris megalotis*	MHNG-1983.070	MHNG	French Guiana	yes	KX381305
C2221	*Noctilio albiventris*	MHNG-1991.014	MHNG	French Guiana	yes	KX381306
C2223	*Noctilio albiventris*	MHNG-1991.016	MHNG	French Guiana	yes	KX381307
C2479	*Trachops cirrhosus*	MHNG-1990.006	MHNG	French Guiana	yes	KX381359
C2511	*Lionycteris spurrelli*	MHNG-1990.011	MHNG	French Guiana	yes	KX381368
C2588	*Diclidurus scutatus*	V-3445	ISEM	French Guiana	yes	KX381370
M1166	*Zygodontomys brevicauda*	MHNG-1975.016	MHNG	French Guiana	yes	KX381437
M1167	*Zygodontomys brevicauda*	MHNG-1974.054	MHNG	French Guiana	yes	KX381438
M1344	*Zygodontomys brevicauda*	MHNG-1991.030	MHNG	French Guiana	yes	KX381443
M1347	*Zygodontomys brevicauda*	MHNG-1975.015	MHNG	French Guiana	yes	KX381444
M1386	*Oecomys bicolor*	MHNG-1996.027	MHNG	French Guiana	yes	KX381448
M1400	*Holochilus sciureus*	MHNG-1996.024	MHNG	French Guiana	yes	KX381450
M1461	*Cryptonanus sp.*	V-3382	ISEM	French Guiana	yes	KX381458
M1462	*Rattus rattus*	MHNG-1991.024	MHNG	French Guiana	yes	KX381459
M1469	*Nectomys rattus*	MHNG-1991.013	MHNG	French Guiana	yes	KX381460
M1525	*Cyclopes didactylus*	V-3456	ISEM	French Guiana	no	KX381462
M1571	*Cyclopes didactylus*	V-3457	ISEM	French Guiana	yes	KX381464
M1664	*Bradypus tridactylus*	V-3450	ISEM	French Guiana	yes	KX381471
M168	*Marmosops pinheiroi*	MHNG-1990.037	MHNG	French Guiana	yes	KX381474
M1748	*Gracilinanus emiliae*	MHNG-1999.001	MHNG	French Guiana	yes	KX381486
M30	*Marmosa demerarae*	MHNG-1996.020	MHNG	French Guiana	yes	KX381531
M31	*Marmosa murina*	MHNG-1979.057	MHNG	French Guiana	yes	KX381532
M333	*Proechimys guyannensis*	MHNG-1984.034	MHNG	French Guiana	yes	KX381535
M336	*Oecomys rutilus*	MHNG-1990.024	MHNG	French Guiana	yes	KX381536
M343	*Oecomys auyantepui*	MHNG-1990.028	MHNG	French Guiana	yes	KX381538
M368	*Proechimys guyannensis*	MHNG-1984.038	MHNG	French Guiana	yes	KX381545
M370	*Hylaeamys yunganus*	MHNG-1991.005	MHNG	French Guiana	yes	KX381546
M376	*Proechimys guyannensis*	MHNG-1984.044	MHNG	French Guiana	yes	KX381547
M381	*Gracilinanus emiliae*	MHNG-1990.032	MHNG	French Guiana	yes	KX381549
M409	*Proechimys guyannensis*	MHNG-1984.020	MHNG	French Guiana	yes	KX381550
M499	*Oecomys auyantepui*	MHNG-1990.034	MHNG	French Guiana	yes	KX381557
M558	*Neacomys paracou*	MHNG-1990.031	MHNG	French Guiana	yes	KX381566
M830	*Marmosa murina*	MHNG-1996.017	MHNG	French Guiana	yes	KX381587
M900	*Hyladelphys kalinowskii*	MHNG-1990.025	MHNG	French Guiana	yes	KX381593
M901	*Oecomys rutilus*	MHNG-1991.021	MHNG	French Guiana	yes	KX381594
M902	*Oligoryzomys fulvescens*	MHNG-1990.022	MHNG	French Guiana	yes	KX381595
M904	*Monodelphis touan*	MHNG-1990.021	MHNG	French Guiana	no	KX381596
M911	*Oligoryzomys fulvescens*	MHNG-1990.038	MHNG	French Guiana	yes	KX381598
M913	*Oligoryzomys fulvescens*	MHNG-1990.039	MHNG	French Guiana	yes	KX381599
M915	*Oligoryzomys fulvescens*	MHNG-1990.049	MHNG	French Guiana	yes	KX381600
M916	*Holochilus sciureus*	MHNG-1990.036	MHNG	French Guiana	yes	KX381601
M936	*Philander opossum*	MNHN-M&O-2013-1960	MNHN	French Guiana	no	KX381605
M938	*Nectomys rattus*	MNHN-M&O-2013-1949	MNHN	French Guiana	yes	KX381606
M939	*Philander opossum*	MNHN-M&O-2013-1953	MNHN	French Guiana	yes	KX381607
M940	*Didelphis marsupialis*	MNHN-M&O-2013-1954	MNHN	French Guiana	yes	KX381608
M942	*Nectomys rattus*	MNHN-M&O-2013-1956	MNHN	French Guiana	yes	KX381609
M943	*Nectomys rattus*	MNHN-M&O-2013-1955	MNHN	French Guiana	yes	KX381610
M945	*Philander opossum*	MNHN-M&O-2013-1959	MNHN	French Guiana	yes	KX381611
M946	*Philander opossum*	MNHN-M&O-2013-1957	MNHN	French Guiana	yes	KX381612
M966	*Mus musculus*	MNHN-M&O-2013-1939	MNHN	French Guiana	yes	KX381616
M968	*Mus musculus*	MNHN-M&O-2013-1940	MNHN	French Guiana	yes	KX381617
M969	*Mus musculus*	MNHN-M&O-2013-1941	MNHN	French Guiana	yes	KX381618
M972	*Mus musculus*	MNHN-M&O-2013-1942	MNHN	French Guiana	yes	KX381619
M983	*Oecomys bicolor*	MHNG-1991.018	MHNG	French Guiana	yes	KX381621

**MNHN= Museum National d'Histoire Naturelle (Paris, France); ROM= Royal Ontario Museum (Toronto, Canada); ISEM= Institut des Science de L'Evolution de Montpellier (France); AMNH= American Museum of National History (New York, USA); MSB= Museum of Southwestern Biology (Albuquerque, USA); ECOBIOS= Etudes en environnement et bioacoustique, Cayenne, French Guiana; MHNG=Museum d'Histoire Naturelle de Genève (Switzerland); EBRG=Estación Biológica Rancho Grande, Maracay, Venezuela*

Appendix 2: Bash script used for the analysis of sequencing data with the OBITOOLS

```
##############################################################################
##Bash script used for the bioinformatic treatment of the sequencing data##
##with the OBITools package (see the complete OBITools documentation at  ##
##(http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12428/pdf).      ##
##############################################################################
# Alignment and merging of paired-end reads using illuminapairedend.
# The program assigns an alignment score to each resulting sequence
# based on the phred quality scores and the length of the aligned regions.

illuminapairedend --fasta-output -r readsR1.fastq readsR2.fastq \
> readsR1R2.fasta

# Reads assignment using ngsfilter. The program requires a table
# providing the information regarding the primer pair and the tag
# combination used for each sample (see the OBITools documentation for more
# details). This step will add an attribute to each sequence containing the
# name of the corresponding sample. Other information such as the
# scientific name of the taxa can be added.

ngsfilter -t ngsfilter.tab -e 2 --nuc readsR1R2.fasta \
 > readsR1R2_ngsfilt.fasta

# Removal of low quality reads using obigrep (alignment scores<50,
# containing Ns or shorter than 50bp)

obigrep -s '^[acgt]+$' -l 50 -p 'score>=50' readsR1R2_ngsfilt.fasta \
> readsR1R2_ngsfilt_lowqual.fasta

# Dereplication of the sequences using obiuniq (regroups every identical
# reads assigned to the same sample into one sequence and keeps the
# coverage information)

obiuniq -c sample readsR1R2_ngsfilt_lowqual.fasta \
> readsR1R2_ngsfilt_lowqual_derep.fasta

# Selecting the majority sequence for each sample using obiselect

obiselect -n 1 -c sample -f count -M readsR1R2_ngsfilt_lowqual_derep.fasta
\ > readsR1R2_ngsfilt_lowqual_derep_maj.fasta

# convert NCBI taxonomic database (release 197) in ecopcr format using
# obiconvert

obiconvert --genbank -t taxdump/ --ecopcrDB-output=ncbi_r197

# Add all taxa included in the reference database into the ncbi taxonomic
# database using obitaxonomy. The reference database in fasta format must
# be first annotated with an attribute containing the name of the taxon and
# an attribute containing the taxonomic path of the sample. The taxonomy
# can be further modified using obitaxonomy (command see he Obitools manual # for
further detail concerning the use of obitaxonomy).

obitaxonomy -d ncbi_r197 -p path -k taxonomy \
-F readsR1R2_ngsfilt_lowqual_derep_maj.fasta

# Add taxids in the attributes of the sequences using obiaddtaxids. The
# program requires an attribute in the sequence containing the scientific
# name of the sample as well as an ecopcr formatted taxonomic database

obiaddtaxids -k taxonomy -d ncbi \
readsR1R2_ngsfilt_lowqual_derep_maj.fasta > mam_12S_refDB.fasta

#The resulting reference database is ready to be used for taxonomic
#assignations of samples with ecotag
```

Appendix 3: Dipteran blood meal analyses procedure and results

Sampling and laboratory procedures

Sand flies and mosquitoes were collected CDC light traps in forest plots in the region of Iracoubo in French Guyana (Counami). After each night, insects caught in the traps were killed by freezing. Blood-fed mosquito and sand fly females were sorted and kept in individual microcentrifuge tubes with absolute 95% ethanol. We extracted DNA from each engorged female using a without-boiling Chelex protocol. The 12S-V5 marker was then amplified, sequenced and taxonomically assigned using the procedure presented in the main document.

Results

Thirty specimens were collected (including for mosquitoes and 26 sand flies). The amplification and sequencing of the 12S-V5 marker was successful for 27 of them, and allowed the identification of eight mammal species belonging to five distinct orders (see table below).

Sample iD	Diptera	Best match in ref. DB	Blood meal identification
CBM_1	mosquito	0.98989899	*Dasypus novemcinctus*
CBM_2	mosquito	0.98989899	*Dasypus novemcinctus*
CBM_3	mosquito	0.98989899	*Dasypus novemcinctus*
CBM_4	mosquito	0.98989899	*Metachirus nudicaudatus*
PBM_1	sand fly	0.98989899	*Dasypus novemcinctus*
PBM_2	sand fly	0.98989899	*Dasypus novemcinctus*
PBM_3	sand fly	0.98989899	*Dasypus novemcinctus*
PBM_4	sand fly	0.98989899	*Dasypus novemcinctus*
PBM_5	sand fly	0.99	*Alouatta macconnelli*
PBM_6	sand fly	0.99	*Alouatta macconnelli*
PBM_7	sand fly	0.99	*Alouatta macconnelli*
PBM_8	sand fly	0.99	*Alouatta macconnelli*
PBM_9	sand fly	0.98989899	*Dasypus novemcinctus*
PBM_10	sand fly	0.989795918	*Dasyprocta leporina*
PBM_11	sand fly	0.99	*Caluromys philander*
PBM_12	sand fly	0.989690722	*Coendou melanurus*
PBM_13	sand fly	0.989690722	*Coendou melanurus*
PBM_14	sand fly	0.989690722	*Coendou melanurus*
PBM_15	sand fly	0.989690722	*Coendou melanurus*
PBM_16	sand fly	0.990196078	*Nasua nasua*
PBM_17	sand fly	0.989690722	*Coendou melanurus*
PBM_18	sand fly	0.98	*Caluromys philander*
PBM_19	sand fly	0.98989899	*Metachirus nudicaudatus*
PBM_20	sand fly	0.98989899	*Dasypus kappleri*
PBM_21	sand fly	0.99	*Alouatta macconnelli*
PBM_22	sand fly	0.990196078	*Nasua nasua*
PBM_23	sand fly	0.99	*Alouatta macconnelli*

ARTICLE 2:
iDNA SCREENING: DISEASE VECTORS AS VERTEBRATE SAMPLERS

Arthur Kocher[1,2], Benoit De Thoisy[3,4], François Catzeflis[5], Sophie Valiere[6], Anne-Laure Bañuls[2] and Jérôme Murienne[1*].

Submitted

[1]CNRS, Université Toulouse III Paul Sabatier, ENFA; UMR5174 EDB (Laboratoire Evolution et Diversité Biologique), Toulouse, France; [2]UMR MIVEGEC (IRD 224 – CNRS 5290 – Université de Montpellier), 911 Avenue Agropolis, F34394 Montpellier, France; [3]Institut Pasteur de la Guyane, 23 avenue Pasteur, 97300 Cayenne, French Guiana; [4]Association Kwata, 16 avenue Pasteur, 97300 Cayenne, French Guiana; [5]Institut des Sciences de l'Evolution, Case Courrier 064, CNRS UMR-5554, Université Montpellier-2, Place E. Bataillon, F-34095 Montpellier, France; [6]GeT–PlaGe, Genotoul, INRA Auzeville, 31326 Castanet-Tolosan, France.

ABSTRACT

In the current context of global change and human-induced biodiversity decline, there is an urgent need for developing sampling approaches able to accurately describe the state of biodiversity. Traditional surveys of the vertebrate fauna involve time-consuming and skill-demanding field methods. Recently, the use of DNA derived from invertebrate parasites (leeches and blowflies) was suggested as a new tool for vertebrate diversity assessment. Blood meal analyses of arthropod disease vectors have long been performed to describe their feeding behaviour, for epidemiological purposes. On the other hand, this important existing expertise has not yet been applied to investigate vertebrate fauna *per se*. Here, we evaluate the usefulness of hematophagous dipterans as vertebrate samplers. Blood-fed sand flies and mosquitoes were collected in Amazonian forest sites, and analysed using high-throughput sequencing of short mitochondrial markers. Blood meal identifications highlighted contrasting ecological features and feeding behaviour among dipteran species, which allowed unveiling arboreal and terrestrial mammals of various body size, as well as birds, lizards and amphibians. Our results also indicate lower vertebrate diversity in sites undergoing higher levels of human-induced perturbation. This suggests that, even with moderate sampling efforts, disease vector blood meals may represent a good proxy for vertebrate communities, opening up promising perspectives for biodiversity monitoring and eco-epidemiology.

BACKGROUND

The current rate of species disappearance is sometimes considered a sixth "mass extinction" crisis [1]. Rapid modifications of biological communities, at both global and local scales, raise concerns as to the concomitant alteration of essential ecosystem functions and services [2]. Developing an indicator system to monitor the state and dynamics of biodiversity through time has been a major challenge to reaching the 2020 Aichi Targets defined by Parties to the United Nations Convention on Biological Diversity. A set of Essential Biodiversity Variables has been recently defined [3], among which taxonomic diversity of communities is relevant to several Aichi Targets. However, there are still uncertainties regarding the economical and practical feasibility of developing monitoring systems at global and local scales.

Vertebrate community composition is frequently used as an indicator for the state of entire biota, because it is relatively well known, occupies a wide variety of ecological niches, provides key ecosystem functions, and contains species that are highly sensitive to human-induced disturbances [1,4–6]. Vertebrate surveys have been relying on time-consuming and expertise-demanding observational methods such as line transects, track counts or live trapping [7–10]. Technological progress has allowed the development of alternative approaches, including remote detection through camera trapping [11] or the sequencing of environmental DNA [12–14]. Recently, the use of DNA derived from invertebrate parasites (iDNA) was suggested as a new tool for vertebrate diversity assessment. Preliminary studies conducted on leeches [15] and carrion flies [16,17] gave promising results but the method has not yet been applied to large datasets.

Host identification from blood meals of disease transmitting arthropods has long been performed for estimating their feeding preference [18–20] because it is an important epidemiological parameter for vector-borne diseases [21–23]. On the other hand, despite their worldwide distribution, the numerous methods available for their collection and analyses, and the good quality of iDNA they provide (i.e. blood; [22]), disease vectors have not been used to investigate the state of vertebrate diversity *per se*.

In this study, we first investigate the usefulness of dipteran blood meal analyses for vertebrate diversity assessment. In particular, we evaluate the applicability of this approach for between-site comparisons. The study was conducted in the Amazonian territory of French Guiana. This region is more than 80% covered by primary lowland rainforests and as such harbours one of the largest continuous rainforest block that benefits from a relatively

favourable conservation status [25]. It thus represents one of the few remaining 'natural laboratories' where undisturbed biodiversity can be studied over large spatial scales. However, demographic pressure and illegal gold mining activities constitute rising threats to biodiversity in part of the territory [6,26,27]. Blood-fed sand flies (Psychodidae: Phlebotominae) and mosquitoes (Culicidae) were collected in forest sites undergoing contrasting levels of anthropogenic pressure. Dipteran and blood meal identifications were performed by PCR amplification and high-throughput sequencing of short mitochondrial markers. The diversity of vertebrates observed through blood meals was then compared between sites after rarefying samples down to equal size and dipteran species composition.

MATERIALS AND METHODS

Sampling

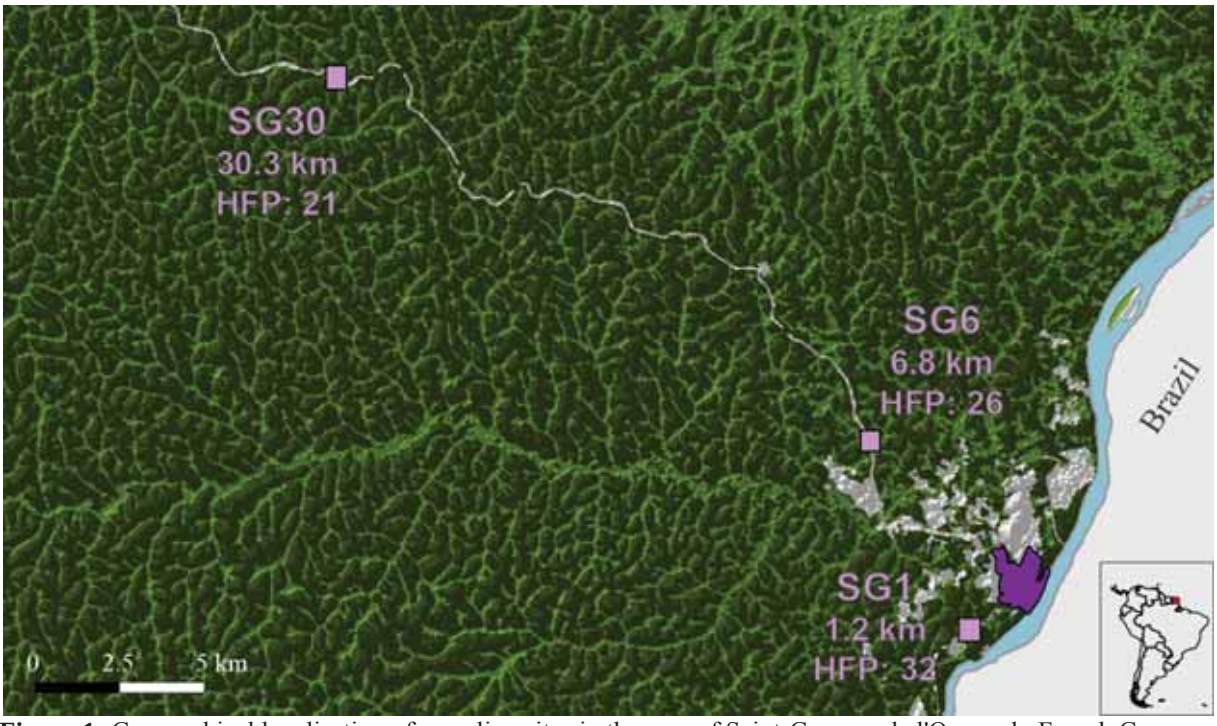


Figure 1: Geographical localisation of sampling sites in the area of Saint-Georges de l'Oyapock, French Guyana. Human footprint index values (HFP, [6]) and distance from town are indicated. Forested areas are represented in green. Saint-Georges town is displayed in dark purple.

Sampling was performed in the area of Saint-Georges de l'Oyapock in March 2016. Three study sites were selected in order to constitute a gradient of anthropogenic pressure, as defined by the Human FootPrint index (HFP index) which has been shown to be a good predictor of vertebrate diversity [6,28]. In particular, hunting pressure was expected to increase with proximity to Saint-Georges town. In order to limit environmental variability, the sites were chosen within the same geomorphological landscape harbouring a highly-dissected

plateau [29] and the same forest type [30]. Site “SG1”, the most impacted site, was located along the road leading to Brazil, close to the Oyapock river, at 1.2 km from Saint-Georges and exhibited an HFP index of 32. Sites “SG6” and “SG30” were located along the national road 2 (constructed 15 years ago) at 6.8 km and 30.3 km from Saint-Georges and presented an HFP index of 26 and 21 respectively (Figure 1).

Nine CDC light-traps (Center for Disease Control) were set in each site, during four consecutive nights (from 6pm to 6am) for a total of 108 trap-nights. After each night, insects contained in the traps were killed by freezing. Blood-fed mosquito and sand fly females were sorted and kept in individual microcentrifuge tubes with 95% ethanol.

DNA amplification and sequencing

We extracted DNA from each engorged female using a without-boiling Chelex protocol [31]. Two PCR were then performed: the first to amplify the Ins16S_1 marker (Ins16S_1-F: TRRGACGAGAAGACCCTATA; Ins16S_1-R: TCTTAATCCAACATCGAGGTC; [32]), which allows species-level identifications of mosquitoes and sand flies ([46] and Dataset S3); the second to amplify the 12S-V5 marker (12S-V5-F: TAGAACAGGCTCCTCTAG; 12S-V5-R: TTAGATACCCCACTATGC; [34]), that was shown to provide accurate identifications of mammals [35]. The 12S-V5 marker was designed for metabarcoding of vertebrates. It thus presents several interesting properties for blood-meal analyses: (i) wide taxonomic coverage (sand flies and mosquitoes may feed on a large variety of vertebrates), (ii) good taxonomic resolution and (iii) very short size (*c.* 100 bp), which allows the detection of highly degraded DNA [13].

The PCR protocol was the same for both fragments. Amplification was performed in 25 µL mixtures containing 3 µL of DNA template, 12 µL of AmpliTaq Gold PCR Master Mix® (5U.µL^{-1}; Applied Biosystems, Foster City, CA, USA), 2.5 µL of each primer (5 µM), and nuclease-free water (Promega, Madison, WI, USA). The PCR mixture was denatured at 95°C (10 min) and followed by 35 cycles of 30s at 95 °C, 30s at 50°C and 30s at 72 °C, completed at 72 °C for 10 mins. Tags of eight base pairs with at least five differences between them were added at the 5’ end of each primer to enable the sequencing of the multiple PCR products in a single sequencing run [36].

PCR products were pooled and sent for library construction and sequencing to the GeT-PlaGe core facilities of Genotoul (Toulouse, France). Samples were diluted in ultrapure water. A volume of 130 µL containing 3 µg of DNA was purified using the HighPrep PCR system (Magbio Genomics, Gaithersburg, MD, USA) and used for library construction with the

Illumina NEXTflex PCR-Free DNA sequencing kit following the instructions of the supplier (Bioo Scientific corp., Austin, TX, USA). Purified fragments were end-repaired, A-tailed and ligated to sequencing indexed adapters. The quality of the library was controlled using the Fragment Analyzer (Advanced Analytical, Ames, IA, USA) and quantified by qPCR with the Library Quantification Kit - Illumina Genome Analyzer-SYBR Fast Universal (CliniSciences, Nanterre, France). The library was pooled with that of other projects and loaded onto the Illumina MiSeq cartridge according to the manufacturer instructions. The quality of the run was checked internally using PhiX. Quality filtering was performed by the Consensus Assessment of Sequence and Variation (CASAVA) pipeline. The sequencing data was stored on the NG6 platform [37] and all computations were performed on the computer cluster of the Genotoul bioinformatic platform (Toulouse, France).

Bioinformatic treatment and taxonomic assignments

The sequencing data was analysed using the `OBITOOLS` package [38], as described previously [33]. Paired-end reads were aligned, merged and then assigned to their corresponding sample based on the tagged primer sequences with two mismatches allowed. Low quality reads were removed and reads were then dereplicated. Taxonomic assignment of sequences was performed with `ecotag` (part or the `OBITOOLS` package). We used reference libraries for Amazonian sand flies [33], mammals [35] and mosquitoes (Dataset S3) that we further complemented with sequences from GenBank using `ecoPCR` [34]. When a sequence had less than 97% identity with its best match in the reference database, we considered the taxonomic assignment at the generic rank or above (as suggested by [47]). Since human presence on the sampling sites was likely scarce and not observed during the sampling period, human sequences identified in blood meals were regarded as laboratory contamination and discarded as a conservative measure. In each sample, only the most abundant sequence was considered (*i.e.* we did not consider potential mixed blood-meals; see *Results and discussion* for further explanations).

When dipterans or vertebrates could not be identified at the species level, we defined molecular taxonomic units (MOTUs) within taxa. We used the Poisson Tree Process [39] as implemented in `mPTP` [40]. The method seeks to classify the branches of a phylogenetic tree into two processes: within species (corresponding to a coalescence process) and between species (corresponding to a speciation process). Because the method uses a phylogenetic tree, we first performed a phylogenetic analysis. Sequences were aligned using `muscle` [41] and a

Maximum Likelihood analysis was performed in RAxML v.8 [42], with the GTR+ Gamma substitution model.

Blood meals diversity analysis

We had to consider two analytical issues when comparing the diversity of dipteran blood meals found between sites. First, the number of blood-fed individuals differed among sites, leading to the well-known "sampling" problem of diversity analysis, referring to the fact that the observed number of species is sensitive to sample size [43]. Secondly, each hematophagous species may exhibit distinct feeding preferences, and their proportions in the samples varied between sites. Hence, the diversity of vertebrates observed through the blood meals is prone to site-specific bias, precluding the use of classical rarefaction analysis or the direct comparison of diversity estimators.

We thus simulated rarefied blood meals pools by randomly drawing, without replacement, a number of individuals from each dipteran species (or MOTU) corresponding to the lowest occurrence of the given species among sites. Using this procedure, the obtained subsamples have identical dipteran composition and can be compared. For each site, we simulated 10 000 rarefied communities in order to estimate the mean frequencies of abundance ranks and Hill numbers of diversity [44]. Sand flies of the same genus appeared to exhibit similar feeding preferences (*i.e. Nyssomyia* feeding mostly on arboreal vertebrates and especially on xenarthrans, *Psychodopygus* feeding mainly on armadillos and *Evandromyia* feeding on terrestrial rodents). Therefore, we grouped sand flies at the generic level in order to increase the maximum base size of rarefied samples. These analyses can be reproduced using data and script available in the Supplementary material.

RESULTS AND DISCUSSION

Dipteran sampling and identifications

In total, 209 blood-fed dipterans were collected, including 99 mosquitoes and 110 sand flies. 77, 76 and 56 specimens were collected in SG30, SG6 and SG1 respectively. The insect Ins16S_1 marker [32] was successfully amplified and sequenced for 94.9% of mosquitoes and 84.5% of sand flies, allowing the identification of 187 specimens (average sequencing depth=6465x). Eight mosquito and eighteen sand fly species (or MOTU) were found (Table 1, Supplementary material).

Blood meal analyses

The vertebrate 12S-V5 marker [34] was successfully amplified and sequenced for 94.9% of mosquitoes and 66.4% of sand flies, allowing taxonomic assignment of 167 blood meals (average sequencing depth=2166x). The lower success rate obtained with sand flies is likely explained by the smaller volume of their blood meals, which is an important factor for successful iDNA amplification [24]. A total of 22 vertebrate MOTUs were found, of which a majority (nineteen) was identified at the species level (Table 1, Supplementary material). The observed diversity encompassed terrestrial and arboreal mammals of various size, as well as birds, lizards and amphibians (Figure 2).

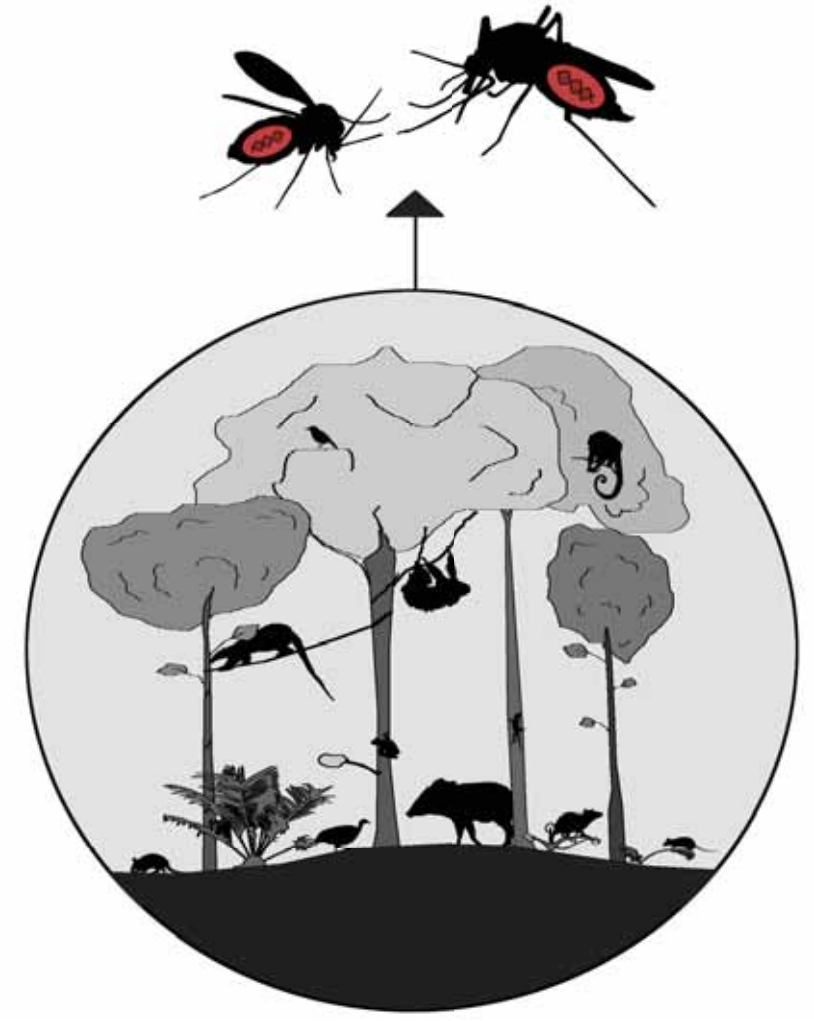


Figure 2: Schematization of the methodology employed in this study. Dipteran blood meal analyses allowed identifying arboreal and terrestrial mammals of various body size as well as birds, lizards and amphibians. This would typically require the implementation of multiple observational methods, wide taxonomic expertise and considerable fieldwork.

The largely dominant mosquito MOTU (*Culex* MOTU 1) seemed to exhibit an opportunistic behaviour, since it had fed on ten mammal species belonging to four different orders, as well as on two bird species (Table 1). *Culex* MOTU 4 was found to feed on two lizards and a frog. All other mosquito species had fed on birds, but were represented by few specimens. For some sand flies species, our observations were consistent with existing knowledge: the canopy-dwelling *Nyssomyia umbratilis* feeding mainly on arboreal mammals including a large proportion of xenarthrans [45,46] and the ground-dwelling *Bichromomyia flaviscutellata* feeding on small rodents [47,48]. In particular, these known vectors of *Leishmania guyanensis* and *L. amazonensis*, respectively, were found to have fed on the main known respective *Leishmania* reservoirs in the region (*Choloepus didactylus* and *Proechimys cuvieri*; [49]). For other species, this study provides the first blood meal data. Five *Psychodopygus* species (*Ps. amazonensis*, *Ps. ayrozai*, *Ps. hirsuta*, *Ps. s. maripaensis* and *Ps.* MOTU 1) fed almost exclusively on armadillos. The nine-banded armadillo (*Da. novemcinctus*) is the known reservoir of *L. naiffi* [50], and *Ps. ayrozay* and *Ps. s. maripaensis* are suspected vectors of the same parasite [51,52]. The association between sand flies and armadillos has been known for a long time, as it can be attested by the amerindian (*Tupi*) name for sand fly, "tatuquira" (litteraly armadillo-fly; [53]). We here report the first direct evidence of this vector-host interaction. The strong observed feeding preference of *Psychodopygus* sand flies for armadillos may explain the relatively low number of human cases caused by *L. naiffi*, as previously suspected, *Ps. ayrozai* being considered as a non-anthropophilic species [54]. Finally, sand flies of the genus *Evandromyia*, for which no diet data was available, were observed to feed mainly on terrestrial rodents. Of anecdotal but surprising mention is the detection of the rodent *Isothrix sinnamariensis* (the brush-tailed rat, weighing ca. 200 g) in a sand fly blood meal. This rare echimyid was previously know from only three localities in French Guiana, one locality in Guyana, and one in Suriname [55].

Several studies have reported the detection of multiple blood meal sources in single hematophagous dipterans [56,57], or the possibility to amplify vertebrate host DNA from specimens that are not visibly blood-fed [58]. In our case, several vertebrate species were identified in 68% of the specimens, with up to 6 and for a mean of 1.95 sequences per individual. However, only considering the majority sequence from visibly blood-fed specimens appeared to us as a necessary conservative measure. PCR-based methods are extremely sensitive to laboratory contaminations, and the use of high-throughput sequencing leads to frequent inter-sample "leakage" due to tag-switching events [59,60]. Of important note, although considering the possibility for mixed blood meal sources would have nearly

double the size of our dataset, it would have conducted to the detection of only one additional vertebrate species (*Potos flavus*, without considering obvious laboratory contaminants that may not be present in the sampling region). In our opinion, this is strong evidence that most of vertebrate identifications we could have considered to originate from mixed blood meals were rather laboratory artefacts. We argue that results obtained from not visibly blood-fed specimens, or the detection of multiple blood meal sources based on DNA amplification methods should be taken with precaution if no proper negative controls (such as strictly phytophagous dipteran males collected in the same traps) were used. This is all the more true for studies reporting the detection of human DNA in arthropods, since it is a frequent laboratory contaminant.

Table 1: Identifications of blood meals in sand flies and mosquitoes collected the area of Saint-Georges de l'Oyapock in French Guiana. Only the 152 specimens for which both dipteran species and blood meals were successfully identified are included.

Group	Species	Nb	Blood meals
Phlebotominae	*Bichromomyia flaviscutellata*	4	*Echimys chrysurus* (1); *Proechimys cuvieri* (2); *Pr.* MOTU 1 (1)
	Evandromyia brachyphalla	2	*Pr. cuvieri* (1); *Pr. guyannensis* (1)
	Ev. infraspinosa	1	*Dasyprocta leporina* (1)
	Ev. sericea	1	*Cuniculus paca* (1)
	Ev. walkeri	1	*Isothrix sinnamariensis* (1)
	Nyssomyia MOTU 1	3	*Choloepus didactylus* (2); *Tamandua tetradactyla* (1)
	Ny. umbratilis	12	*Alouatta macconelli* (1); *Ch. didactylus* (7); *Coendou melanurus* (2); *Pecari tajacu* (1); *Ta. tetradactyla* (1)
	Pintomyia damascenoi	1	*Ta. tetradactyla* (1)
	Psychodopygus amazonensis	3	*Dasypus kappleri* (1); *Du. novemcinctus* (2)
	Ps. ayrozai	17	*Du. kappleri* (2); *Du. novemcinctus* (14); *Pe. tajacu* (1)
	Ps. claustrei	1	*Dr. leporina* (1)
	Ps. hirsuta	5	*Du. novemcinctus* (5)
	Ps. s. maripaensis	8	*Du. novemcinctus* (7); *Pe. tajacu* (1)
	Ps. MOTU 1	5	*Du. novemcinctus* (5)
	Sciopemyia sordellii	1	*Pr. cuvieri* (1)
Culicidae	*Aedes serratus*	1	*Thamnophilus* MOTU 1 (1)
	Culex mollis	1	*Tinamus major* (1)
	Culex MOTU 1	78	*Ch. didactylus* (3); *Co. melanurus* (1); *Cu. paca* (6); *D. leporina* (2); *Du. novemcinctus* (1); *M. nudicaudatus* (3); *Pe. tajacu* (1); *Pr. cuvieri* (37); *Pr. guyannensis* (1); *Pr.* MOTU 1 (4); *Th. nigrocinereus* (16); *Th.* MOTU 1 (3)
	Culex MOTU 2	1	*Th. nigrocinereus* (1)
	Culex MOTU 3	2	*Th. nigrocinereus* (2)
	Culex MOTU 4	3	*Kentropyx calcarata* (1); *Osteocephalus MOTU* (1); *Polychrus marmoratus* (1)
	Wyeomyia ypsipola	1	*Ti. major* (1)

Diversity comparisons

The percentage of blood meal pairwise identity was higher within trap-nights than within sites, for both mosquitoes (87.6% versus 32.4%; X^2=347.68; $p<10^{-15}$) and sand flies (53.5% versus 31.6%; X^2=7.91; p<0.005), indicating that insects collected in the same trap and during the same night were likely to have fed on the same individual or group of conspecifics. Therefore, blood meal data was converted into occurrences by trap-night for each dipteran species (or genus for sand flies, see Material and Methods) before diversity analysis.

Decreasing biodiversity is a typical signature of human-induced disturbance [61,62] and was expected along our sampling sites. Taking raw blood meal data, we observed a marked decrease in vertebrate species richness with increasing HFP and proximity to town (seventeen, eleven and seven species in SG30, SG6 and SG1 respectively, Figure 3).

However, since different numbers of specimens were collected in each site, with varying proportions of dipteran species characterized by specific feeding habits, direct comparisons of observed blood meal richness are inevitably biased. Rarefaction is a commonly employed procedure to account for sampling size in diversity analyses [43]. Here, we rarefied our sample in order to estimate Hill numbers of diversity [44] based on blood meals originating from identical pools of dipterans.

The rarefied samples included 13 randomly drawn specimens: four *Culex* MOTU 1, one *Evandromyia*, three *Nyssomyia* and five *Psychodopygus*. Resulting estimated Hill numbers of order zero, one and two (*i.e.* species richness, exponential of the Shannon-Wiener index and Simpson index) all decreased with increasing HFP index and proximity to town (Figure 2). Of note, some species that are typically rare or absent in highly disturbed habitat were found only in site SG30 or SG6 (*e.g. Alouatta maconnelli*, *Metachirus nudicaudatus*, *Pecari tajacu*, *Tamandua tetradactyla*; *Tinamus major*; [6, 10]). This suggests that dipteran blood meals may provide a relevant proxy for vertebrate diversity comparisons, even with moderate sampling effort.

Dipteran species collected in this study exhibited a variety of ecological features and complementarity feeding behaviour. Therefore, the analyses of the whole dipteran community, rather than specific species, allowed the detection of a broad spectrum of vertebrate hosts. We argue that future studies should develop integrated sampling protocols that are able to maximise the diversity of iDNA-carrying arthropods collected on study sites. However, such a community-based approach also comes with drawbacks, since between-site

comparisons are prone to biases related to arthropod community compositions. In this study, we adopted a conservative approach in this regard, but the development of specific analytical tools is another main challenge to be addressed in future iDNA studies.

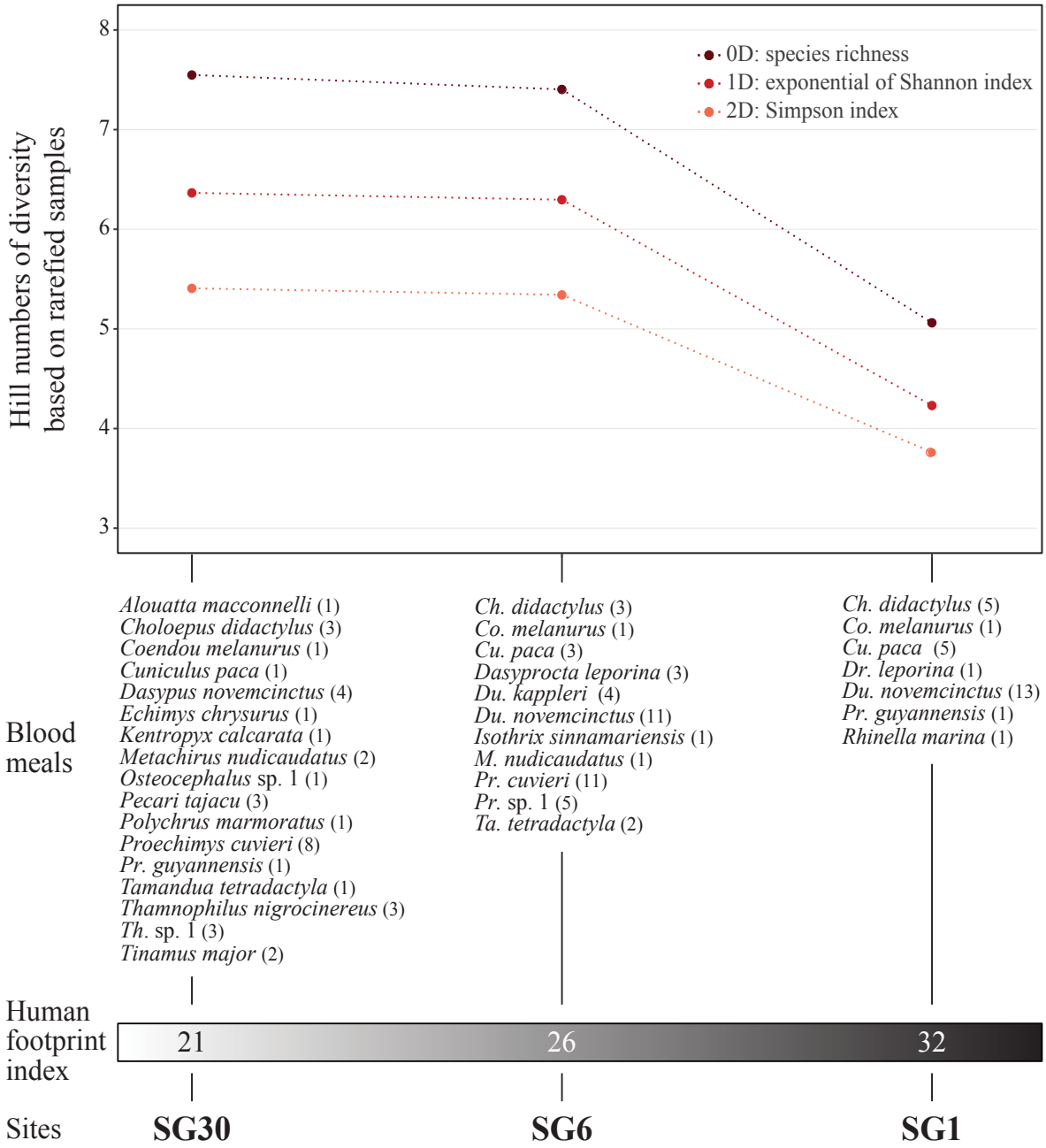


Figure 3: Vertebrates identified in dipteran blood meals (numbers in brackets indicate occurrences per trap-night) and estimations of Hill numbers of diversity (qD) on rarefied blood meal pools among sites, along an anthropogenic pressure gradient defined by the human footprint index [6].

CONCLUSIONS

This work demonstrates the usefulness of dipteran blood meal analyses as a single source of information to screen for the presence of a wide range of vertebrates. Our study also

suggests that the approach may be suitable for highlighting variation in vertebrate community composition between sampling sites. Knowing the large expertise existing for the study of disease vectors, these results set the stage for promising perspectives.

Arthropods themselves are already frequently used as biodiversity indicators [63,64] and blood meal identifications could easily be complemented with insect community analyses in order to extend the scope of such studies. Furthermore, our results provide precious contributions to the knowledge of disease vector feeding habits. In recent years, questions have been raised about the impact of biodiversity changes on the transmission risk of zoonotic infectious diseases [65]. In combination with DNA-based pathogen detection, the present approach could open up new avenues for both biodiversity assessment and eco-epidemiology.

DATA AVAILABILITY

All data and scripts are available in the Supplementary material.

AUTHORS' CONTRIBUTIONS

AK, A-L B, BT and JM designed research; AK, JM, FC and BT performed field work; AK and SV performed laboratory work; AK analysed the data; AK wrote the paper and all authors contributed to its improvement.

FUNDING

This work was supported by “Investissement d’Avenir” grants managed by Agence Nationale de la Recherche (LABEX CEBA: ANR-10-LABX-25-01; LABEX TULIP: ANR-10-LABX-41, ANR-11-IDEX-0002-02), Observatoire Homme-Mileu Oyapock (USR 3456, within the frame of LABEX DRIIHM) as well as the Inkermann fund hosted at the Fondation de France.

ACKNOWLEDGMENTS

We would like to thank Pierre-Michel Forget and François Feer for their precious help in the design of this study; Antoine Fouquet (LEEISA CNRS), Romain Girod and Pascal Gaborit (Institut Pasteur de la Guyane) for logistical support; Stanislas Talaga for providing us the mosquito reference database and Fabian Fischer for his kind revision of the English language.

REFERENCES

1. Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2015 Accelerated modern human–induced species losses: Entering the sixth mass extinction. Sci. Adv. 1, e1400253. (doi:10.1126/sciadv.1400253)
2. Cardinale BJ et al. 2012 Biodiversity loss and its impact on humanity. Nature 486, 59–67. (doi:10.1038/nature11148)
3. Pereira HM *et al.* 2013 Essential Biodiversity Variables. *Science* **339**, 277–278. (doi:10.1126/science.1229931)
4. Ceballos G, Ehrlich PR. 2002 Mammal Population Losses and the Extinction Crisis. *Science* **296**, 904–907. (doi:10.1126/science.1069349)
5. Collen B, Loh J, Whitmee S, McRAE L, Amin R, Baillie JEM. 2009 Monitoring Change in Vertebrate Abundance: the Living Planet Index. *Conserv. Biol.* **23**, 317–327. (doi:10.1111/j.1523-1739.2008.01117.x)
6. de Thoisy B, Richard-Hansen C, Goguillon B, Joubert P, Obstancias J, Winterton P, Brosse S. 2010 Rapid evaluation of threats to biodiversity: human footprint score and large vertebrate species responses in French Guiana. *Biodivers. Conserv.* **19**, 1567–1584. (doi:10.1007/s10531-010-9787-z)
7. Voss RS, Emmons LH. 1996 *Mammalian diversity in Neotropical lowland rainforests: a preliminary assessment*. New York: American Museum of Natural History.
8. Smallwood KS, Fitzhugh EL. 1995 A track count for estimating mountain lion Felis concolor californica population trend. *Biol. Conserv.* **71**, 251–259. (doi:10.1016/0006-3207(94)00034-N)
9. Vieira EM, Monteiro-Filho ELA. 2003 Vertical stratification of small mammals in the Atlantic rain forest of south-eastern Brazil. *J. Trop. Ecol.* **19**, 501–507. (doi:10.1017/S0266467403003559)
10. de Thoisy B, Brosse S, Dubois MA. 2008 Assessment of large-vertebrate species richness and relative abundance in Neotropical forest using line-transect censuses: what is the minimal effort required? *Biodivers. Conserv.* **17**, 2627–2644. (doi:10.1007/s10531-008-9337-0)
11. Rovero F, Martin E, Rosa M, Ahumada JA, Spitale D. 2014 Estimating species richness and modelling habitat preferences of tropical forest mammals from camera trap data. *PLOS ONE* **9**, e103300. (doi:10.1371/journal.pone.0103300)
12. Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjær KH, Orlando L, Gilbert MTP, Willerslev E. 2012 Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.* **21**, 1966–1979. (doi:10.1111/j.1365-294X.2011.05261.x)
13. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012 Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* **21**, 2045–2050.
14. Valentini A *et al.* 2016 Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.* **25**, 929–942. (doi:10.1111/mec.13428)
15. Schnell IB, Thomsen PF, Wilkinson N, Rasmussen M, Jensen LRD, Willerslev E, Bertelsen MF, Gilbert MTP. 2012 Screening mammal biodiversity using DNA from leeches. *Curr. Biol.* **22**, R262–R263. (doi:10.1016/j.cub.2012.02.058)
16. Calvignac-Spencer S, Merkel K, Kutzner N, Kühl H, Boesch C, Kappeler PM, Metzger S, Schubert G, Leendertz FH. 2013 Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Mol. Ecol.* **22**, 915–924. (doi:10.1111/mec.12183)
17. Lee P-S, Gan HM, Clements GR, Wilson J-J. 2016 Field calibration of blowfly-derived DNA against traditional methods for assessing mammal diversity in tropical forests. *Genome* **59**, 1008–1022. (doi:10.1139/gen-2015-0193)
18. Tesh RB, Chaniotis BN, Carrera BR, Johnson KM. 1972 Further studies on the natural host preferences of Panamanian phlebotomine sandflies. *Am. J. Epidemiol.* **95**, 88–93.
19. Tempelis CH. 1975 Review article: host-feeding patterns of mosquitoes, with a review of advances in analysis of blood meals by serology. *J. Med. Entomol.* **11**, 635–653.
20. Humair P-F, Douet V, Cadenas FM, Schouls LM, Van De Pol I, Gern L. 2007 Molecular identification of bloodmeal source in ixodes ricinus ticks using 12S rDNA as a genetic marker. *J. Med. Entomol.* **44**, 869–880. (doi:10.1603/0022-2585(2007)44[869:MIOBSI]2.0.CO;2)
21. Simpson JE, Hurtado PJ, Medlock J, Molaei G, Andreadis TG, Galvani AP, Diuk-Wasser MA. 2012 Vector host-feeding preferences drive transmission of multi-host pathogens: West Nile virus as a model system. *Proc. R. Soc. Lond. B Biol. Sci.* **279**, 925–933. (doi:10.1098/rspb.2011.1282)
22. Hamer GL, Chaves LF, Anderson TK, Kitron UD, Brawn JD, Ruiz MO, Loss SR, Walker ED, Goldberg TL. 2011 Fine-scale variation in vector host use and force of infection drive localized patterns of west nile virus transmission. *PLoS ONE* **6**, e23767. (doi:10.1371/journal.pone.0023767)
23. Kent RJ. 2009 Molecular methods for arthropod bloodmeal identification and applications to ecological and vector-borne disease studies. *Mol. Ecol. Resour.* **9**, 4–18. (doi:10.1111/j.1755-0998.2008.02469.x)

24. Calvignac-Spencer S, Leendertz FH, Gilbert MTP, Schubert G. 2013 An invertebrate stomach's view on vertebrate ecology: Certain invertebrates could be used as 'vertebrate samplers' and deliver DNA-based information on many aspects of vertebrate ecology. *BioEssays* **35**, 1004–1013. (doi:10.1002/bies.201300060)
25. Hammond DS. 2005 *Tropical Forests of the Guiana Shield: Ancient Forests in a Modern World*. CABI.
26. de Thoisy B, Renoux F, Julliot C. 2005 Hunting in northern French Guiana and its impact on primate communities. *Oryx* **39**, 149–157. (doi:10.1017/S0030605305000384)
27. Hammond DS, Gond V, de Thoisy B, Forget P-M, DeDijn BPE. 2007 Causes and Consequences of a Tropical Forest Gold Rush in the Guiana Shield, South America. *AMBIO J. Hum. Environ.* **36**, 661–670. (doi:10.1579/0044-7447(2007)36[661:CACOAT]2.0.CO;2)
28. de Thoisy B, Fayad I, Clément L, Barrioz S, Poirier E, Gond V. 2016 Predators, Prey and Habitat Structure: Can Key Conservation Areas and Early Signs of Population Collapse Be Detected in Neotropical Forests? *PLOS ONE* **11**, e0165362. (doi:10.1371/journal.pone.0165362)
29. Guitet S, Cornu J-F, Brunaux O, Betbeder J, Carozza J-M, Richard-Hansen C. 2013 Landform and landscape mapping, French Guiana (South America). *J. Maps* **9**, 325–335.
30. Guitet S, Brunaux O, de Granville J, Richard-Hansen C. 2015 *Catalogue des habitats forestiers de Guyane*. Cayenne: ONF and DEAL.
31. Casquet J, Thebaud C, Gillespie RG. 2012 Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. *Mol. Ecol. Resour.* **12**, 136–141. (doi:10.1111/j.1755-0998.2011.03073.x)
32. Clarke LJ, Soubrier J, Weyrich LS, Cooper A. 2014 Environmental metabarcodes for insects: *in silico* PCR reveals potential for taxonomic bias. *Mol. Ecol. Resour.* **14**, 1160–1170. (doi:10.1111/1755-0998.12265)
33. Kocher A *et al.* 2017 Vector soup: high-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Mol. Ecol. Resour.* **17**, 172–182. (doi:10.1111/1755-0998.12556)
34. Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E. 2011 ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Res.* **39**, 1–11. (doi:10.1093/nar/gkr732)
35. Kocher A, Thoisy BD, Catzeflis F, Huguin M, Valiere S, Zinger L, Bañuls A-L, Murienne J. in press Evaluation of short mitochondrial metabarcodes for the identification of Amazonian mammals. *Methods Ecol. Evol.* (doi:10.1111/2041-210X.12729)
36. Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E. 2007 The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* **2**, e197. (doi:10.1371/journal.pone.0000197)
37. Mariette J, Escudié F, Allias N, Salin G, Noirot C, Thomas S, Klopp C. 2012 NG6: Integrated next generation sequencing storage and processing environment. *BMC Genomics* **13**, 462.
38. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. 2016 obitools: a unix-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* **16**, 176–182.
39. Zhang J, Kapli P, Pavlidis P, Stamatakis A. 2013 A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869–2876.
40. Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T. 2016 Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *bioRxiv* (doi:10.1101/063875)
41. Edgar RC. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
42. Stamatakis A. 2014 RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
43. Gotelli NJ, Colwell RK. 2011 Estimating species richness. *Biol. Divers. Front. Meas. Assess.* **12**, 39–54.
44. Chao A, Gotelli NJ, Hsieh TC, Sander EL, Ma KH, Colwell RK, Ellison AM. 2014 Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67.
45. Le Pont F, Pajot F-X. 1980 Leishmaniose en Guyane Francaise. 1. Etude de l'écologie et du taux d'infection naturelle du vecteur *Lutzomyia* (*Nyssomyia*) *umbratilis* Ward et Fraiha, 1977 en saison sèche. Considérations épidemiologiques. *Cah. ORSTOM Entomol. Medicale Parasitol.* **18**, 359–382.
46. Christensen HA, Arias, JR A, de Vasquez, AM de V, de Freitas, RA de F. 1982 Hosts of sandfly vectors of *Leishmania braziliensis guyanensis* in the central Amazon of Brazil. *Am. J. Trop. Med. Hyg.* **31**, 239–242.
47. Lainson R, Shaw JJ. 1968 Leishmaniasis in Brazil: I. Observations on enzootic rodent leishmaniasis—incrimination of *Lutzomyia flaviscutellata* as the vector in the lower amazonian basin. *Trans. R. Soc. Trop. Med. Hyg.* **62**, 396–405.
48. Tesh RB, Chaniotis BN, Aronson MD, Johnson KM, others. 1971 Natural host preferences of Panamanian phlebotomine sandflies as determined by precipitin test. *Am. J. Trop. Med. Hyg.* **20**, 150–156.

49. Rotureau B. 2006 Ecology of the *Leishmania* species in the Guianan ecoregion complex. *Am. J. Trop. Med. Hyg.* **74**, 81–96.
50. Lainson R, Shaw J. 1988 *Leishmania* (*Viannia*) *naiffi* sp. n., a parasite of the armadillo, *Dasypus novemcinctus* (L.) in Amazonian Brazil. *Ann. Parasitol. Hum. Comp.* **64**, 3–9.
51. De Freitas R, Biancardi C, Castellon E. 1985 Flagellate infections of brazilian sand flies (diptera: psychodidae): isola tion in vitro and biochemical identification of *Endotrypanum* and *Leishmania*. *Am J Trap Med Hyg* **34**, 1098–1108.
52. Fouque F, Gaborit P, Issaly J, Carinci R, Gantier J-C, Ravel C, Dedet J-P. 2007 Phlebotomine sand flies (Diptera: Psychodidae) associated with changing patterns in the transmission of the human cutaneous leishmaniasis in French Guiana. *Mem. Inst. Oswaldo Cruz* **102**, 35–40.
53. Von Ihering R. 1968 *Dicionário dos animais do Brasil*. Brasilia: Edit. Univ. de Brasilia.
54. Lainson R, Shaw JJ. 2010 New world leishmaniasis. *Topley Wilsons Microbiol. Microb. Infect.* (doi:10.1002/9780470688618.taw0182)
55. Lim BK, Catzeflis F. 2014 History of mammalogy in the Guianas. In *Historia de la mastozoologia en Latinoamerica, las Guayanas y el Caribe* (eds J Ortega, L Luis Martinez, DG Tirira), pp. 265–292. Quito and Mexico, DF: Murcielago Blanco and Asociacion Ecuatoriana de Mastozoologia.
56. Sales KG da S, Costa PL, de Morais RCS, Otranto D, Brandão-Filho SP, Cavalcanti M de P, Dantas-Torres F. 2015 Identification of phlebotomine sand fly blood meals by real-time PCR. *Parasit. Vectors* **8**. (doi:10.1186/s13071-015-0840-3)
57. González E, Gállego M, Molina R, Abras A, Alcover MM, Ballart C, Fernández A, Jiménez M. 2015 Identification of blood meals in field captured sand flies by a PCR-RFLP approach based on cytochrome b gene. *Acta Trop.* **152**, 96–102. (doi:10.1016/j.actatropica.2015.08.020)
58. Moreno M, Saavedra MP, Bickersmith SA, Prussing C, Michalski A, Rios CT, Vinetz JM, Conn JE. 2017 Intensive trapping of blood-fed *Anopheles darlingi* in Amazonian Peru reveals unexpectedly high proportions of avian blood-meals. *PLoS Negl. Trop. Dis.* **11**, e0005337. (doi:10.1371/journal.pntd.0005337)
59. Esling P, Lejzerowicz F, Pawlowski J. 2015 Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res.* **43**, 2513–2524. (doi:10.1093/nar/gkv107)
60. Schnell IB, Bohmann K, Gilbert MTP. 2015 Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Mol. Ecol. Resour.* **15**, 1289–1303. (doi:10.1111/1755-0998.12402)
61. Turner IM. 1996 Species Loss in Fragments of Tropical Rain Forest: A Review of the Evidence. *J. Appl. Ecol.* **33**, 200–209. (doi:10.2307/2404743)
62. Brashares JS, Arcese P, Sam MK, Coppolillo PB, Sinclair ARE, Balmford A. 2004 Bushmeat Hunting, Wildlife Declines, and Fish Supply in West Africa. *Science* **306**, 1180–1183. (doi:10.1126/science.1102425)
63. Talaga S, Dézerald O, Carteron A, Leroy C, Carrias J-F, Céréghino R, Dejean A. 2017 Urbanization impacts the taxonomic and functional structure of aquatic macroinvertebrate communities in a small Neotropical city. *Urban Ecosyst.* , 1–9.
64. Dedieu N, Vigouroux R, Cerdan P, Céréghino R. 2015 Invertebrate communities delineate hydro-ecoregions and respond to anthropogenic disturbance in East-Amazonian streams. *Hydrobiologia* **742**, 95–105.
65. Ostfeld RS, Keesing F. 2012 Effects of Host Diversity on Infectious Disease. *Annu. Rev. Ecol. Evol. Syst.* **43**, 157–182. (doi:10.1146/annurev-ecolsys-102710-145022)

APPENDICES

Appendix 1: Sample details and sequencing data

Appendix 1 will be available with the publication of this article

Appendix 2: R script used for the analysis of blood meal diversity

```
##########################################################################
##                                                                      #
##              Dipteran Blood meal diversity analyses                  #
## Kocher et al. 2017                                                   #
##########################################################################
#This script was used to detect autocorrelation of blood meal data within trap
night and to perform rarefaction analyses for diversity comparison between sites

####IMPORT DATA (Dataset S1 in tab-delimited format)####
Blood_meal_data=read.table(file='S1.txt',header=T,quote="",stringsAsFactors =
F,sep='\t')

####BLOOD MEAL SIMILARITY WITHIN TRAP-NIGHT####
#definition of a function used to detect autocorrelation of the data within
trap_night for a set of samples
trap_night_similarity=function(data){
  #computation of the number of pairwise identical blood meals within each sampling
site
  pairwise_id_site=NULL
  for (i in 1:length(unique(data$Site))){
    if (sum(data$Site==unique(data$Site)[i])>1){
      comb=combn(1:sum(data$Site==unique(data$Site)[i]),2);
      data_bis=data[data$Site==unique(data$Site)[i],];
      for (j in 1:ncol(comb)){

pairwise_id_site=c(pairwise_id_site,as.numeric(data_bis$Blood.meal.identification[c
omb[1,j]]==data_bis$Blood.meal.identification[comb[2,j]]))
      }
    }
  }
  #computation of the number of pairwise identical blood meals within each trap-
night
  pairwise_id_trap_night=NULL
  for (i in 1:length(unique(data$Trap.night.ID))){
    if (sum(data$Trap.night.ID==unique(data$Trap.night.ID)[i])>1)
      {
      comb=combn(1:sum(data$Trap.night.ID==unique(data$Trap.night.ID)[i]),2);
      data_bis=data[data$Trap.night.ID==unique(data$Trap.night.ID)[i],];
      for (j in 1:ncol(comb)){

pairwise_id_trap_night=c(pairwise_id_trap_night,as.numeric(data_bis$Blood.meal.iden
tification[comb[1,j]]==data_bis$Blood.meal.identification[comb[2,j]]))
      }
    }
  }
  #return a list containing:
  ##(i) the result of a chi2 test comparing the proportion of identical samples
within sampling site versus within trap-night
  ##(ii) the proportion of identical samples within sampling site versus within
trap-night
  ##(iii) the standard errors of the proportions

return(list(chi2=prop.test(x=c(sum(pairwise_id_site),sum(pairwise_id_trap_night)),n
=c(length(pairwise_id_site),length(pairwise_id_trap_night))),means=c(mean(pairwise_
id_site),mean(pairwise_id_trap_night)),se=c(sqrt(mean(pairwise_id_site)*(1-
mean(pairwise_id_site))/length(pairwise_id_site)),sqrt(mean(pairwise_id_trap_night)
*(1-mean(pairwise_id_trap_night))/length(pairwise_id_trap_night)))))
}
```

```
#use of the fonction on mosquito and sand flies separately, removing missing blood
meal (12S-V5) data
TNS_culicidae=trap_night_similarity(Blood_meal_data[Blood_meal_data$Dipteran.family
=='Culicidae'&Blood_meal_data$Blood.meal.identification!='12S-V5 failed',])
TNS_psychodidae=trap_night_similarity(Blood_meal_data[Blood_meal_data$Dipteran.fami
ly=='Psychodidae'&Blood_meal_data$Blood.meal.identification!='12S-V5 failed',])

#barplot
prop=cbind(TNS_culicidae$means,TNS_psychodidae$means)
colnames(prop)=c("Mosquitoes", "Sand flies")
se=cbind(TNS_culicidae$se,TNS_psychodidae$se)

x=barplot(prop,ylim = c(0,1.2),beside=T,axes=F,ylab = "Proportion of pairwise
identity",col = c("gray40","gray90"))
arrows(x,prop-2*se,x,prop+2*se,angle=90,code = 3,lty = 1,length = 0.1)
axis(side=2,at=seq(0,1,0.2))
axis(side=1,lwd.ticks = 0,labels = F)
legend("topright",inset = 0.05,legend=c("Within sites","Within trap-nights"),fill =
c("gray40","gray90"),bty="n")

####RAREFACTION ANALYSIS####
#definition of a function to estimate Hill numbers of diversity based on rarefied
blood meal pools of equal size and dipteran composition
#data=dataset to analyse
#sites=list of sampling sites to compare
#nrep=number of random replicates to draw from each site for estimating Hill
numbers
rarefy_blood_meals=function(data,sites,nrep){
  #create a variable to group dipteran exhibiting similar feeding habits (referred
to as "trophic species"; we choose the MOTU or species for mosquitoes and the genus
for sand flies)

data$trophic_species[data$Dipteran.family=='Culicidae']=data$Dipteran.identificaton
[data$Dipteran.family=='Culicidae']

data$trophic_species[data$Dipteran.family=='Psychodidae']=data$Dipteran.genus[data$
Dipteran.family=='Psychodidae']
  #factorise variables
  data$trophic_species=as.factor(data$trophic_species)
  data$Blood.meal.identification=as.factor(data$Blood.meal.identification)
  #reduce dataset to occurence by trophic species and

data=unique(data[,c("Site","Trap.night.ID","trophic_species","Blood.meal.identifica
tion")])
  #define the composition of the rarefied blood meal pools as the the minimum
occurence of each species among sampling sites

rarefied_composition=apply(table(data[data$Site%in%sites,c("Site","trophic_species"
)]),MARGIN = 2,FUN=min)
  #prepare an output containing the size and composition of the rarefied blood meal
pool and estimated Hill numbers for each site

result=list(m=sum(rarefied_composition),rarefied.composition=rarefied_composition,H
ill.numbers=data.frame(D0=rep(NA,length(sites)),D1=rep(NA,length(sites)),D2=rep(NA,
length(sites)),row.names = sites))
  #for each sampling site
  for (s in sites){
    tab=data[data$Site==s,]
    #prepare a vector of vertebrate taxa abundance frequencies (see Chao et al.
2014 for details)
    abund_freq=rep(0,sum(rarefied_composition)+1)
    names(abund_freq)=as.character(seq(0,sum(rarefied_composition)))
    #for nrep replicates
    for (i in 1:nrep){
      #draw a random rarefied blood meal pool
      rarefied_pool=rep(0,length(table(tab$Blood.meal.identification)))
      names(rarefied_pool)=names(table(tab$Blood.meal.identification))
      for (i in unique(tab$trophic_species)){
```

```
rarefied_pool=rarefied_pool+table(sample(tab$Blood.meal.identification[tab$trophic_
species==i],size = rarefied_composition[i],replace = FALSE))
      }
      #compute of the abundace frequencies in the random pool and adding in the
vector
      random_abund_freq=table(rarefied_pool)

abund_freq[names(random_abund_freq)]=abund_freq[names(random_abund_freq)]+random_ab
und_freq
    }
    #divide the abundance frequencies vector by the number of replicates to get
mean abundance frequencies
    mean_abund_freq=abund_freq/nrep
    #remove null abundances
    mean_abund_freq=mean_abund_freq[-1]
    #estimate Hill numbers based on mean abundance frequencies (see Chao et al.
2014)
    D0=sum(mean_abund_freq)
    D1=exp(sum(-
as.numeric(names(mean_abund_freq))/length(mean_abund_freq)*log(as.numeric(names(mea
n_abund_freq))/length(mean_abund_freq),exp(1))*mean_abund_freq))

D2=1/sum((as.numeric(names(mean_abund_freq))/length(mean_abund_freq))^2*mean_abund_
freq)
    #fill output with estimated Hill numbers for the given site
    result$Hill.numbers[s,]=c(D0,D1,D2)
  }
  return(result)
}

#use the rarefy_blood_meals function to compare the 3 sampling between them with
10000 replicates. We first remove missing dipteran or blood meal identifications
from the data
rarefy_3sites=rarefy_blood_meals(Blood_meal_data[Blood_meal_data$Dipteran.identific
aton!='Ins16S_1 failed'&Blood_meal_data$Blood.meal.identification!='12S-V5
failed',],c("SG30","SG6","SG1"),nrep = 10000)
#sites can be compared pairwise to get more statistical power (larger rarefied
blood meals)
rarefy_SG30_SG6=rarefy_blood_meals(Blood_meal_data[Blood_meal_data$Dipteran.identif
icaton!='Ins16S_1 failed'&Blood_meal_data$Blood.meal.identification!='12S-V5
failed',],c("SG30","SG6"),nrep = 10000)
rarefy_SG30_SG1=rarefy_blood_meals(Blood_meal_data[Blood_meal_data$Dipteran.identif
icaton!='Ins16S_1 failed'&Blood_meal_data$Blood.meal.identification!='12S-V5
failed',],c("SG30","SG1"),nrep = 10000)
rarefy_SG6_SG1=rarefy_blood_meals(Blood_meal_data[Blood_meal_data$Dipteran.identifi
caton!='Ins16S_1 failed'&Blood_meal_data$Blood.meal.identification!='12S-V5
failed',],c("SG6","SG1"),nrep = 10000)
```

CHAPITRE 4: LEISHMANIOSES ET BIODIVERSITÉ : EXPLORATION DE L'EFFET DE DILUTION

RÉSUMÉ

La finalité de ce travail de thèse était d'étudier l'impact de la perturbation des habitats sur le cycle de transmission des leishmanioses. Plus particulièrement, il s'agit de tester l'hypothèse de l'effet de dilution sur ce modèle d'étude multi-pathogènes, multi-vecteurs et multi-hôtes.

Ce chapitre est composé de deux parties. **Dans un premier article**, nous présentons une méthode d'identification des *Leishmania* basée sur le séquençage de minicercles kinétoplastiques. Les minicercles sont de petits éléments génomiques circulaires caractéristiques des kinetoplastidés. Leur très grand nombre de copies cellulaires, et la présence de blocs conservés dans leur séquence permettant la conception d'amorces de PCR en font une cible de choix pour la détection sensible des parasites. Toutefois, la nature hétérogène des populations des minicercles présentent dans chaque cellule limite l'utilisation de cette région génomique pour l'identification d'espèces avec les techniques classiques. Nous montrons que le séquençage haut-débit d'un court fragment de minicercle amplifié par PCR permet d'effectuer des identifications taxonomiques fiables.

Dans la deuxième partie, nous appliquons la méthode pour estimer l'abondance des phlébotomes infectés le long du gradient d'anthropisation dans la région de Saint-Georges, sur lequel un déclin de diversité à été observé au travers des repas sanguins de diptères (Chapitre 3). Les données générées précédemment sont réutilisées pour montrer que les conditions nécessaires à la manifestation d'un effet de dilution semblent réunies le long du gradient: **(i)** l'abondance des vecteurs ne diminue pas **(ii)** les espèces de mammifères connues pour être réservoir de *Leishmania* sont aussi celles qui persistent sur les sites les plus pauvres en espèces de vertébrés. La prévalence parasitaire observée est globalement très faible, mais quatre espèces de *Leishmania* sont identifiées dans nos échantillons. Conformément à l'hypothèse de l'effet de dilution, la prévalence est maximale sur le site où la diversité de vertébrés est la plus faible. Les différences observées entre les sites ne sont toutefois pas significatives. D'avantage de données sont nécessaires pour tirer des conclusions solides sur les liens existant entre biodiversité et leishmanioses. En particulier, la mise en place de ce type d'étude dans des zones ou la circulation du parasite est plus intense pourrait permettre l'observation de patrons plus marqués.

ARTICLE 1:

ULTRA-SENSITIVE IDENTIFICATION OF NEOTROPICAL *LEISHMANIA* SPP. BASED ON HIGH-THROUGHPUT SEQUENCING OF KINETOPLAST MINICIRCLE AMPLICONS

Arthur Kocher[1,2], Sophie Valière[3], Anne-Laure Bañuls[2] and Jérôme Murienne[1*].

Submitted

[*1]CNRS, Université Toulouse III Paul Sabatier, ENFA; UMR5174 EDB (Laboratoire Evolution et Diversité Biologique), Toulouse, France; [2]UMR MIVEGEC (IRD 224 – CNRS 5290 – Université de Montpellier), 911 Avenue Agropolis, F34394 Montpellier, France; [3]GeT–PlaGe, Genotoul, INRA Auzeville, 31326 Castanet-Tolosan, France.

ABSTRACT

Reliable identifications of *Leishmania* spp. are essential for medical management and eco-epidemiological studies. This is particularly true in the Neotropical region where multiple species of these parasites are frequently found in sympatry. Kinetoplast minicircles are the preferred targets for sensitive *Leishmania* detection with wide taxonomic coverage, because they are present in high copy number and contain conserved sequence blocks in which PCR primers can be designed. On the other hand, the heterogenic nature of minicircle networks has hampered the use of this peculiar genomic region for species identifications. Here, we present a barcoding-like approach based on high-throughput sequencing of short minicircle amplicons that allows reliable identification of *Leishmania* spp. The 120 bp long minicircle conserved region was amplified by PCR from 18 *Leishmania* strains representative of the major species complexes found in the Neotropics. High-throughput sequencing of PCR products enabled recovering significant numbers of distinct minicircle sequences from each strain, reflecting minicircle class diversity. We show that minicircle sequence analysis can provide correct species identification, even when conspecific reference strains are isolated in distant geographic locations. Given current technological developments, we argue that high-throughput sequencing of kDNA minicircles may soon represent a powerful tool for a variety of applications in *Leishmania* research from biological samples with low parasite loads.

INTRODUCTION

Leishmaniases are a group of diseases caused by flagellate parasites of the genus *Leishmania* Ross, 1903 (Kinetoplastida: Trypanosomatidae). They are endemic in approximately 100 countries in Africa, America, Asia, and Europe, and cause more than one million cases and several thousand deaths each year (Alvar *et al.*, 2012). More than 50 *Leishmania* species have been described to date, among which approximately 20 are known to infect humans (Maroli *et al.*, 2013; Akhoundi *et al.*, 2016). These may cause various clinical forms (e.g. localized cutaneous, disseminated cutaneous, mucocutaneous or visceral leishmaniasis) and respond differently to treatment (Copeland and Aronson, 2015). Furthermore, each *Leishmania* species is characterized by a specific transmission cycle involving distinct sand fly vectors and reservoir hosts (Lainson and Shaw, 2010; Maroli *et al.*, 2013; Ready, 2013). In areas where several *Leishmania* species are found in sympatry or when imported cases can occur, reliable identifications of strains are crucial for medical management and eco-epidemiological studies.

For more than 40 years, multilocus enzyme electrophoresis (MLEE) has allowed considerable advances in the delineation and classification of *Leishmania* species (Gardener and Howells, 1972; Kreutzer *et al.*, 1983; Rioux *et al.*, 1990; Thomaz-Soccol *et al.*, 1993). Nevertheless, this technique is time-consuming and relies on constraining steps of parasite isolation and culture. PCR-based methods are increasingly used as alternatives to MLEE for *Leishmania* typing, because they are cost and time-effective and can be performed on small quantities of material. In particular, Restriction Fragment Length Polymorphism (RFLP) analyses (Marfurt *et al.*, 2003; Rotureau, 2006; Koarashi *et al.*, 2016) and direct comparisons of DNA sequences (Zelazny *et al.*, 2005; Marco *et al.*, 2006; de Almeida *et al.*, 2011) have been widely employed in recent years. Various genomic regions have been targeted for molecular identifications of *Leishmania*, in a quest for the best compromise between sensitivity, specificity and taxonomic coverage (Akhoundi *et al.*, in press), but no real consensus has been reached in this regard.

Leishmania parasites, like all kinetoplastids, are characterised by a unique mitochondrion containing a remarkably large and dense genome: the kinetoplast. Kinetoplast DNA (kDNA) represents 10-20% of total cellular DNA (Simpson, 1987) and is composed of two types of circular molecules interlinked in a concatenated network: maxicircles and minicircles. Maxicircles are *c.* 20 kb long, present in a few tens of copies, and are analogous to the mitochondrial genome of other eukaryotes. Minicircles are *c.* 800 bp long, present in

several thousands of copies. They encode for guide RNAs (gRNAs) involved in the maturation of maxicircle's messenger RNAs through an RNA editing mechanism (Read *et al.*, 2015). Due to their extremely high copy number, minicircles are ideal targets for highly sensitive detection of *Leishmania*. Furthermore, the presence of well-conserved sequence blocks (CSBs) in their replication origin (Jensen and Englund, 2012) allows the design of PCR primers with wide taxonomic coverage (e.g. Noyes *et al.*, 1998; Ceccarelli *et al.*, 2014).

On the other hand, the use of minicircle sequences for *Leishmania* species identification is difficult. The kDNA network present in each parasitic cell encompasses a set of distinct minicircle versions, referred to as classes, encoding for different gRNAs (Hajduk and Ochsenreiter, 2010). Except in the CSBs, high sequence variability is found among minicircle classes (Brewster and Barker, 2002), resulting in difficulties for their analysis (de Oliveira Ramos Pereira and Brandão, 2013). Furthermore, there are redundant gRNAs that encode the same editing information, and the number and identity of minicircle classes varies greatly between species and even between strains of the same species (Simpson, 1997; Gao *et al.*, 2001). These features make classical DNA-based identification method difficult to perform. Thus, in many studies, minicircle-based PCR assays are employed as highly sensitive screening tests for *Leishmania* detection, while other genomic regions are targeted in a second step for the identification of positive samples (Richini-Pereira *et al.*, 2014; Cássia-Pires *et al.*, 2014; Berzunza-Cruz *et al.*, 2015; Pereira Júnior *et al.*, 2015).

Here, we present a novel approach based on high-throughput sequencing of short minicircle fragments amplified by PCR, which aims at combining sensitive detection and reliable identification of all *Leishmania* spp. The method is evaluated for a use in the Neotropical region, where a high diversity of *Leishmania* species may be found in sympatry (Lainson and Shaw, 2010).

MATERIAL AND METHOD

Leishmania strains and DNA extraction

Eighteen *Leishmania* strains were selected, representative of both *Leishmania* subgenera and of all major species complexes found in the Neotropics (Table 1). They belonged to six species (as identified by MLEE or PCR-RFLP; Simon *et al.*, 2010): *L.* (*Viannia*) *braziliensis*, *L.* (*V.*) *guyanensis*, *L.* (*V.*) *lainsoni*, *L.* (*V.*) *naiffi*, *L.* (*Leishmania*) *amazonensis* and *L.* (*L.*) *infantum*. When possible, we included several strains isolated in distant locations for each *Leishmania* species. *Leishmania* parasites were cultured as

previously described (Faye *et al.*, 2010). DNA was extracted from culture pellets using the Pure Link Genomic DNA Kit (Life Technologies, Saint-Aubin, France).

Table 1: List of the strains used in this study

Short code	WHO code	Species	Country	Lab./ref.
2007-24	MHOM/GF/2007/2007-24	*L. amazonensis*	French Guiana	(Simon *et al.* 2010)
LV78	RAT/BR/72/LV78	*L. amazonensis*	Brazil	CRB-L*
2006-9	MHOM/GF/2006/2006-9	*L. braziliensis*	French Guiana	(Simon *et al.* 2010)
LC1412	MHOM/PE/91/LCI412	*L. braziliensis*	Peru	(Dujardin *et al.* 1995)
LH699	MHOM/PE/00/LH699	*L. braziliensis*	Peru	(Victoir *et al.* 1998)
LH754	MHOM/PE/89/LH754	*L. braziliensis*	Peru	(Dujardin *et al.* 1995)
M2904	MHOM/BR/75/M2904	*L. braziliensis*	Brazil	CRB-L
2006-3	MHOM/GF/2006/2006-3	*L. guyanensis*	French Guiana	(Simon *et al.* 2010)
2007-1	MHOM/GF/2007/2007-1	*L. guyanensis*	French Guiana	(Simon *et al.* 2010)
M5378	MHOM/BR/78/M5378	*L. guyanensis*	Brazil	CRB-L
HZ2008	MHOM/GF/2008/HZ2008	*L. guyanensis*	French Guiana	CRB-L
TD2008	MHOM/GF/2008/TD2008	*L. guyanensis*	French Guiana	CRB-L
ITMAP263	MHOM/MA/67/ITMAP263	*L. infantum*	Marocco	CRB-L
LEM417	MHOM/DZ/1982/LIPA59	*L. infantum*	Algeria	CRB-L
2006-40	MHOM/GF/2006/2006-40	*L. lainsoni*	French Guiana	(Simon *et al.* 2010)
LC2288	MHOM/PE/91/LC2288	*L. lainsoni*	Peru	(Bastrenta *et al.* 2002)
LEM2204	MDAS/BR/79/M5533	*L. naiffi*	Brazil	CRB-L
LEM5108	MHOM/GF/2005/LEM5108	*L. naiffi*	French Guiana	CRB-L

*Centre de Ressources Biologiques des Leishmania, Monptellier, France.

PCR amplification

PCR amplification targeted the *c.* 120 bp portion of kDNA minicircle known as the minicircle conserved region. We used PCR primers located within CSB1 and CSB3, similar to previously published ones (*e.g.* Roque *et al.*, 2010; Weirather *et al.*, 2011; Fig. 1), but further adjusted based on a large set of complete minicircle sequences (leishmini-F: 5'-GGKAGGGGCGTTCTGC-3'; leishmini-R: 5'-STATWTTACACCAACCCC-3'). Our aim was to ensure that all minicircle classes will be amplified in every *Leishmania* spp.

Amplification was performed in 20 µL mixtures containing 2 µL of DNA template, 10 µL of AmpliTaq Gold PCR Master Mix® (5U.µL^{-1}; Applied Biosystems, Foster City, CA, USA), 2.5 µL of each primer (5 µM), and nuclease-free water (Promega, Madison, WI, USA). The PCR mixture was denatured at 95°C (10 min) and followed by 35 cycles of 30s at 95°C, 30s at 60°C and 15s at 72°C, completed at 72°C for 5 mins. To enable the sequencing of multiple PCR products in a single high-throughput sequencing run, tags of eight base pairs with at least five differences between them were added at the 5' end of each primer (Binladen

et al., 2007). In order to evaluate the frequency of tag-switching events, which may result in some sequences being assigned to the wrong sample (Schnell *et al.*, 2015), we only used one out of two available tag combinations, as suggested by (Esling *et al.*, 2015). PCR sensitivity was checked on serial dilution of *Leishmania* DNA followed by electrophoresis and visualization of PCR product in a 2% agarose gel containing SYBR Safe stain (Invitrogen) under UV light.

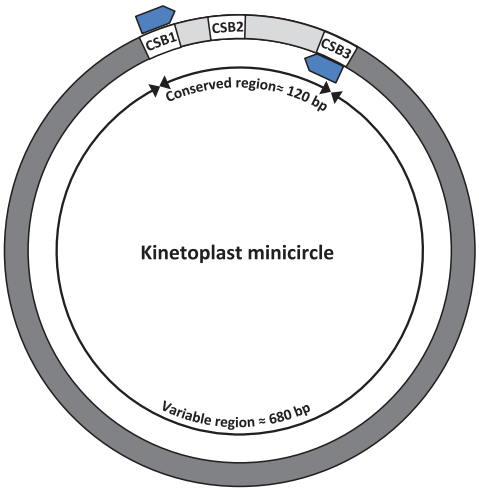


Figure 1: Structural organization of *Leishmania* kinetoplast minicircles. The three conserved sequence blocks (CSBs) are indicated in white boxes. Blue arrows indicate the binding sites of leishmini primers used to amplify a *c.* 80-bp-long fragment within the conserved region.

High-throughput sequencing of minicircle amplicons

PCR products obtained from the 18 *Leishmania* strains were pooled and sent to the GeT-PlaGe core facilities of GenoToul (Toulouse, France) for library construction and sequencing on an Illumina Hiseq 3000 platform (Illumina, San Diego, CA, USA). Quality filtering was performed with the Consensus Assessment of Sequence and Variation (CASAVA) pipeline. Sequence data were stored on the NG6 platform (Mariette *et al.*, 2012).

Sequencing reads were analysed with the OBITOOLS (Boyer *et al.*, 2016), as described elsewhere (Kocher *et al.*, in press, 2017). Pair-end reads were aligned and merged, taking into account the phred quality scores for consensus construction. Merged reads were assigned to a given sample based on the primer tags with two mismatches allowed. Low quality reads (exhibiting an alignment scores < 50, containing Ns or shorter than 50bp) were discarded and then dereplicated in each sample. To limit errors arising from tag-switching events, we discarded all sequences that had coverage below that of the most abundant sequence found with an unexpected tag combination. The bash script used for these bioinformatic steps is available in the Supplementary Material.

Minicircle sequence analysis and use for species identification

Because multiple alignments of minicircle sequences are problematic (de Oliveira Ramos Pereira and Brandão, 2013), we performed pairwise alignments of each sequence pair based on longest common subsequences (Needleman-Wunsch alignment with a score of one for matches and zero for mismatches and gaps). Raw distances were then used to generate a neighbour-joining (NJ) tree with the R package 'ape' (Saitou and Nei, 1987; Paradis *et al.*, 2004; R Core team, 2014), in order to visualise clustering patterns among *Leishmania* strains and species. Distance statistics were computed using the R package 'spider' (Brown *et al.*, 2012).

Our aim was to use a barcoding-like strategy to perform *Leishmania* species identifications based on minicircle sequences. For each *Leishmania* strain, several minicircle classes were expected to be amplified and recovered by high-throughput sequencing. Therefore, taxonomic assignments of *Leishmania* spp. required the use of a multilocus approach. However, contrary to multilocus barcoding, where several markers are sequenced separately (Fazekas *et al.*, 2008; Dufour *et al.*, 2016), we did not have *a priori* knowledge on the identity of each minicircle fragments retrieved. It was thus not possible to perform sequence alignment for each minicircle classes independently. Additionally, the set of minicircle classes may vary between strains, even within the same *Leishmania* species (Simpson, 1997; Gao *et al.*, 2001). Thus, there was no guaranty that each minicircle sequence recovered from a strain would find a homologue among those of other conspecific references. Furthermore, some minicircle classes may be species-specific, while others are shared between species (Brewster and Barker, 2002). Therefore, we had to consider that only a fraction of sequences would be informative for taxonomic assignments.

Our rationale was the following: when comparing a set of minicircle sequences from an unidentified *Leishmania* strain to that of reference strains, each may either (i) match closely with one or several homologous references all belonging to the same species (*i.e.* specific match), (ii) match closely with several homologous references from various species (*i.e.* unspecific match), or (iii) not match closely with any available reference (*i.e.* no homologue, Fig. 2). Eventually, the taxonomic assignment of the strain should be based on the subset of minicircles for which specific matches were found.

In practice, for a given unidentified strain, we used the ecotag program (included in the OBITOOLS) to perform taxonomic assignment of each minicircle sequence. ecotag includes a "last common ancestor algorithm" that allows to distinguish between specific and unspecific

match. The strain was then taxonomically assigned according to the most frequent specific match, after weighting by sequence abundances (Fig. 2). The proportion of specific matches supporting the resulting identification was used as an identification score.

To evaluate the accuracy of the method, we performed a "leave one out" testing. Each strain (i.e. the full set of minicircle sequences) was consecutively considered unknown and identified using the method described above with the remaining strains as references. Bash and R scripts used to perform taxonomic assignments of strains and to compute identification scores are available in the Supplementary Material.

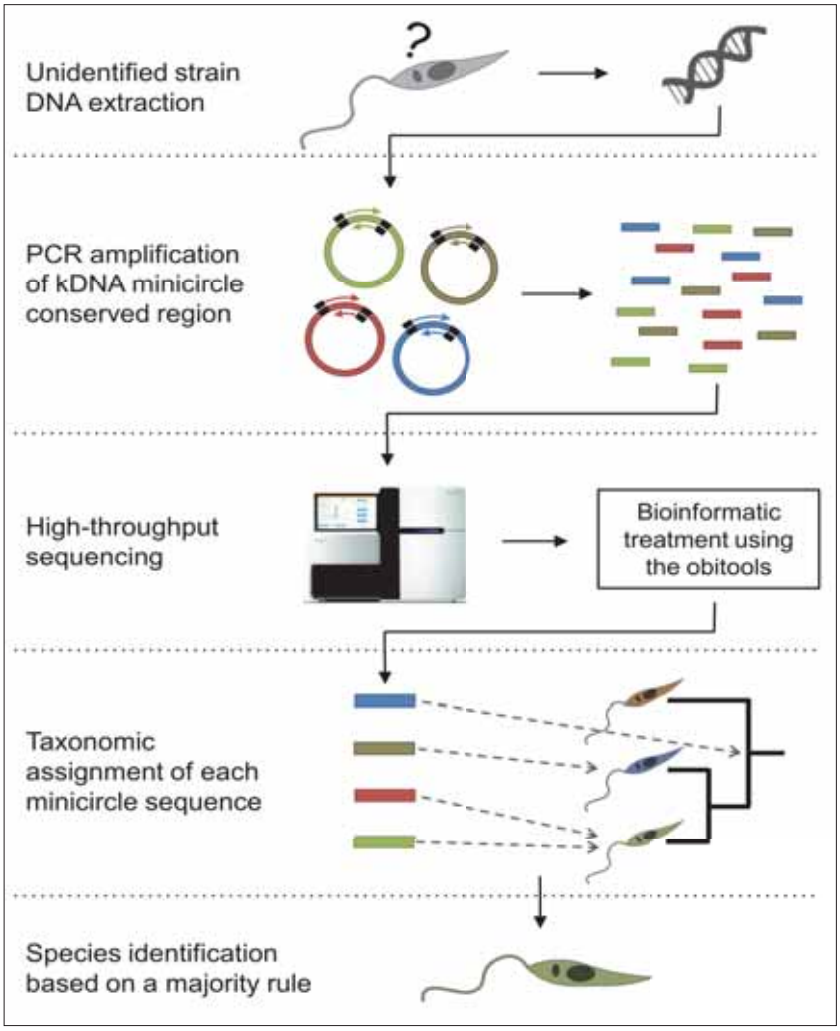


Figure 2: Schematization of the pipeline used for the identification of *Leishmania* spp. based on the high-throughput sequencing of *c.* 80-bp-long kDNA minicircle amplicons

RESULTS

High-throughput sequencing of minicircle amplicons

PCR amplification of kDNA minicircles allowed the detection of all *Leishmania* species down to 1 fg of template DNA (*i.e.* below 10^{-2} parasite equivalent; see Supplementary material). High-throughput sequencing of PCR products allowed obtaining a total of 10,347,227 pairs of 150 bp long reads. After paired-end merging, quality filtering removed 0.23% of reads. 670,890 unique sequences remained after read dereplication. The most abundant sequence assigned to a non-used tag combination had a coverage of 1,172, which was used as a threshold to remove sequences potentially originating from tag-switching events. Remaining sequences had an average read coverage of 4,380. Eventually, the number of minicircle sequences retrieved ranged from 10 to 84 depending on the strain, for an average of 37.2 (Table 3; GenBank acc. KY699529 - KY700198).

Minicircle sequence analysis and use for species identification

Based on pairwise alignment of minicircle sequences, average intraspecific and interspecific raw genetic distances were 11.5% and 26.4% respectively. The highest mean intraspecific distance was found in *L. amazonensis* (18.9%; Table 2), while the lowest were found in *L. naiffi, L. guyanensis* and *L. braziliensis* (10.3, 10.4% and 10.6% respectively).

Table 2: Number of distinct minicircle sequences retrieved and genetic distance statistics for each Leishmania species

Species	Number of sequences	Mean intraspecific distance (%)	Mean intrastrain distance (%)	Mean interstrain distance (%)
L. amazonensis	62	18.9	18.6	19.2
L. braziliensis	133	10.6	10.7	10.5
L. guyanensis	196	10.4	10.5	10.4
L. infantum	94	12.6	12.6	12.5
L. lainsoni	97	14.7	14.5	14.8
L. naiffi	88	10.3	10.4	10.3

In the NJ tree, minicircle sequences of each strain did not form single monophyletic clusters, but were rather distributed in several clusters comprising sequences of other strains, generally belonging to the same species (Fig. 3). This pattern was also highlighted by very similar intrastrain and interstrain genetic distance within each species (Table 2). At the

species level, several patterns were observed. Almost all sequences were grouped in single species-specific clusters for *L. amazonensis*, *L. infantum* and *L. lainsoni*. For the three other species (*L. naiffi*, *L. braziliensis* and *L. guyanensis*), minicircle sequences formed several interleaved clusters. Some of these were well delimited and species-specific while others included a mix of species (for *L. braziliensis* and *L. guyanensis* especially).

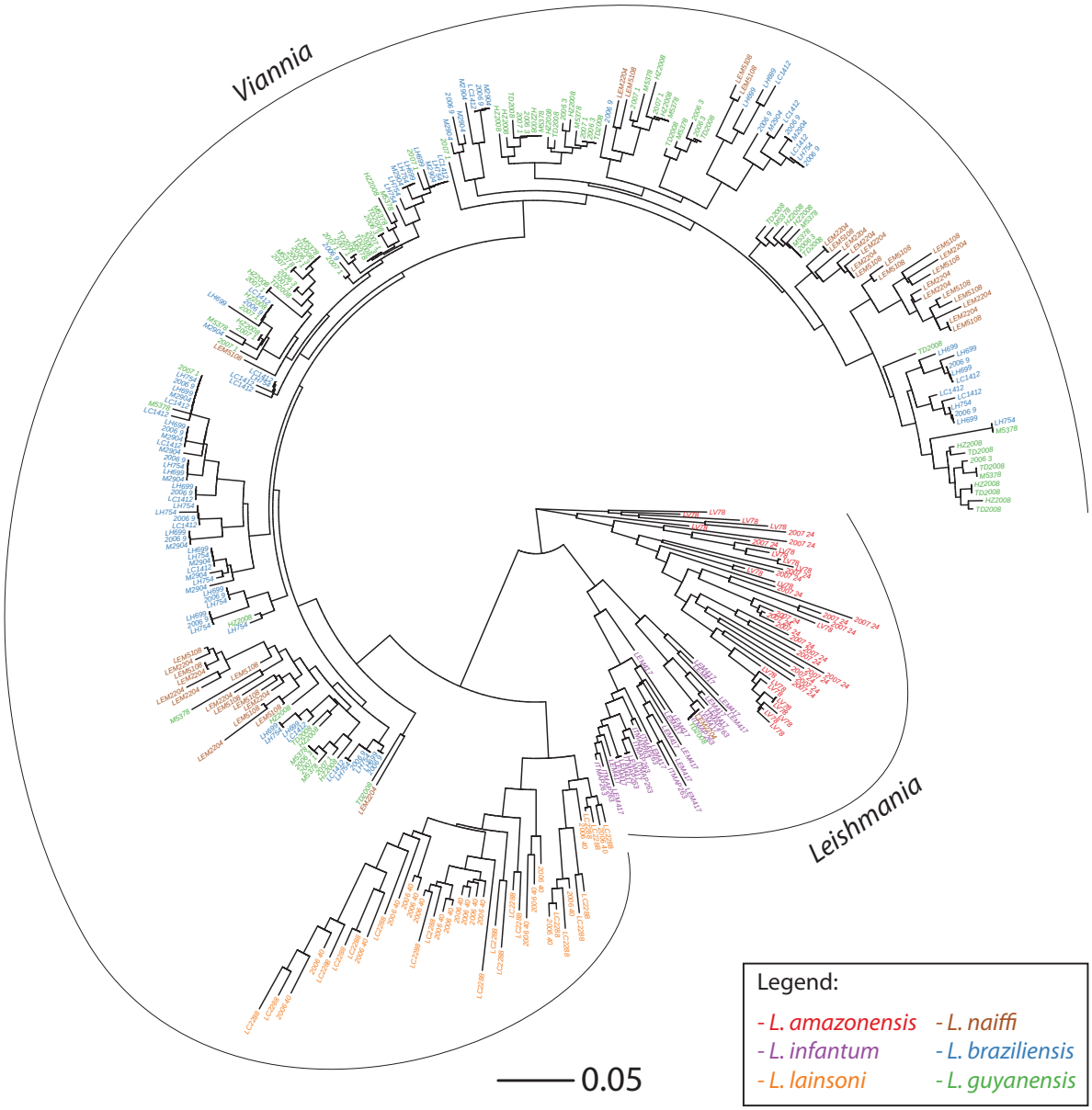


Figure 3: Neighbor-joining tree based on pairwise alignments of *c.* 80-bp-long kDNA minicircle fragments. Only the 20 most abundant sequences amplified for each strain are included. The colors refer to *Leishmania* species. Each sequence is referenced with its strain ID.

The leave-one-out procedure showed that, using the method presented here, all strains would have been correctly identified if considered unknown (Table 3). On average, for each strain, 84.8% of minicircle sequences matched on species-specific references, and these were indicative of the correct *Leishmania* species in 97.85% of cases. Most strains (10/18) were

identified with an identification score of 100%, including all *L. amazonensis*, *L. infantum*, *L. lainsoni* and *L. naiffi* strains. Some minicircle sequences of *L. braziliensis* identified as *L. guyanensis* and *vice versa*, which resulted in identification scores below 100%. However, most of these were identified with a score above 95%. The most ambiguous identifications were that of strains M5378 and HZ2008 (scores=90.4% and 88.4% respectively), but these were still correctly identified as *L. guyanensis*.

Table 3: Taxonomic identification of the strain using a leave-one-out procedure.

Strain	Species	Nb. of seq.	Proportion of taxonomic assignations (%)*							Identification	Score (%)
			LA	LB	LG	LI	LL	LN	NS		
2007_24	*L. amazonensis*		100	0	0	0	0	0	0	*L. amazonensis*	100
LV78	*L. amazonensis*	41	100	0	0	0	0	0	0	*L. amazonensis*	100
2006_9	*L. braziliensis*	21	0	66.4	0	0	0	1.6	32	*L. braziliensis*	97.6
LC1412	*L. braziliensis*	26	0	78.2	2.7	0	0	0	19.1	*L. braziliensis*	96.7
LH699	*L. braziliensis*	30	0	75.5	3	0	0	0	21.5	*L. braziliensis*	96.2
LH754	*L. braziliensis*	37	0	75	1.5	0	0	0	23.5	*L. braziliensis*	98
M2904	*L. braziliensis*	23	0	74.4	3.4	0	0	0	22.2	*L. braziliensis*	95.6
2006_3	*L. guyanensis*	17	0	0	100	0	0	0	0	*L. guyanensis*	100
2007_1	*L. guyanensis*	12	0	1.5	92.7	0	0	0	5.8	*L. guyanensis*	98.4
M5378	*L. guyanensis*	40	0	5.1	61.3	0	0	1.4	32.3	*L. guyanensis*	90.4
HZ2008	*L. guyanensis*	44	0	10.7	81.7	0	0	0	7.5	*L. guyanensis*	88.4
TD2008	*L. guyanensis*	58	0	0	94.3	0	0	0	5.7	*L. guyanensis*	100
ITMAP263	*L. infantum*	42	0	0	0	56	0	0	44	*L. infantum*	100
LEM417	*L. infantum*	10	0	0	0	69.7	0	0	30.3	*L. infantum*	100
LC2288	*L. lainsoni*	84	0	0	0	0	85.8	0	14.2	*L. lainsoni*	100
2006_40	*L. lainsoni*	51	0	0	0	0	100	0	0	*L. lainsoni*	100
LEM2204	*L. naiffi*	46	0	0	0	0	0	93.7	6.3	*L. naiffi*	100
LEM5108	*L. naiffi*	50	0	0	0	0	0	90.8	9.2	*L. naiffi*	100

*Proportion of sequences reads assigned to each species; LA=*L. amazonensis*, LB=*L. braziliensis*, LG=*L. guyanensis*, LI=*L. infantum*, LL=*L. lainsoni*, LN= *L. naiffi*, NS= Not specifically assigned

DISCUSSION

The kDNA minicircle conserved region is known to be an ideal target for detection of all *Leishmania spp.* with high sensitivity (Akhoundi *et al.*, in press). This was further confirmed by our results, since PCR amplification was successful for the six *Leishmania* species belonging to both *Viannia* and *Leishmania* subgenera, down to extremely low DNA concentration. On the other hand, we would like to stress that other non-*Leishmania* kinetoplastid could probably be detected with the same PCR assay due to the high

conservation of CSBs, which should not be used for specific diagnosis without confirmation by sequencing.

Despite the usefulness of kDNA minicircles for *Leishmania* detection, classical DNA-based identification methods may hardly be applicable for these peculiar genomic molecules. One of the reasons is that several distinct minicircle classes coexist in each parasitic cell. Even if all minicircles can be amplified simultaneously with the same PCR primers located in the CSBs, the comprehensive characterization of minicircle population use to require isolation and cloning of kDNA or PCR products prior to sequencing (Lee *et al.*, 1993; Brewster and Barker, 2002; Telleria *et al.*, 2006; Rodrigues *et al.*, 2013). High-throughput sequencing allows bypassing these laborious tasks and will likely contribute to fulfil important methodological gaps in kinetoplastid genomics. By sequencing of single PCR products, we were able to easily retrieve significant numbers of distinct minicircle fragments in each strain, reflecting the diversity of minicircle classes.

Another difficulty comes with the variability in the number and identity of minicircle classes found among strains, even within the same species. Therefore, we used a barcoding-like strategy that does not rely on *a priori* knowledge of minicircle homology. Our results show that, using our method, every strain would have been correctly identified, even when conspecific reference strains were isolated from very distant geographic locations. This highlights the robustness of the approach for the identification of all major species complexes found in the New World. Hence, the method may be employed as a single assay for sensitive *Leishmania* detection and identification in this region characterized by high *Leishmania* diversity. To go further, the development of similar strategies targeting the minicircle variable region may allow the typing of strains at finer resolution. Indeed, kDNA minicircle fingerprinting has long been known as a valuable tool for describing and tracking *Leishmania* genetic diversity, with a discriminatory power superior to that of MLEE (Angelici *et al.*, 1989; Botilde *et al.*, 2006).

One limitation of the approach is that it may be unable to identify several *Leishmania* spp. in the same sample (*e.g.* in case of coinfection, or when analysing pools of sand flies, Kocher *et al.*, 2017). Indeed, it may not be possible to conclude on the presence of distinct species when conflicting signals arise from minicircle sequences, since this may happen within single strains. The constitution of more comprehensive and local minicircle reference databases will probably narrow the spectrum of species or strain-specific minicircle sequences, which could help resolve this issue. Another drawback is that high-throughput sequencing remains relatively costly unless large batches of samples are analyzed, and that

several days or weeks may be required from the preparation of the samples to the bioinformatic treatment of the data. Such an approach is therefore not adapted for all applications yet. However, given the rapid evolution of these technologies, high-throughput sequencing will likely be widely accessible in a near future.

Even if this study did not aimed at providing phylogenetic information, our analyses revealed patterns of sequence clustering that are remarkably congruent with current hypothesis of *Leishmania* relationships and timings of speciation (Harkins *et al.*, 2016). In particular, the observation of the NJ tree seems to indicate that the presence of shared or homologue minicircle classes between species tend to decrease when the given species are more distantly related. More extensive studies based on similar high-throughput sequencing approaches could probably provide precious insights into the evolutionary aspects of kinetoplast minicircle networks. Additionally, the method presented here could be adapted to acquire knowledge in fundamental biological processes of these parasites. Minicircle analyses in *Trypanosoma brucei* and *T. cruzi* allowed to evidence clear signatures of intraspecific recombination (Gibson *et al.*, 1997). Population structure in *Leishmania* still needs to be further detailed, especially as to the occurrence of genetic exchanges (Rougeron *et al.*, in press). High-throughput sequencing of minicircle amplicons may represent a powerful tool in this regard.

CONCLUSIONS

High-throughput sequencing opens new avenues for the study of kDNA minicircles. Here, we showed the efficiency of this technology to recover fragments from various minicircle classes simultaneously amplified by PCR. Despite known difficulties for minicircle sequence analysis, we developed a barcoding-like strategy that proved to be robust for all major *Leishmania* species complexes found in the New World, while taking advantage of minicircle properties for ultra-sensitive PCR detection. In the future, the development of similar approaches could be useful for describing and tracking *Leishmania* diversity at infraspecific levels or for studying *Leishmania* population structure from samples containing low concentration of parasite DNA. Given the rapid evolution and increasing accessibility of these technologies, high-throughput sequencing of kDNA minicircles may soon represent a powerful tool for a variety of common applications in *Leishmania* research.

SUPPLEMENTARY MATERIAL

Bash and R scripts used for the bioinformatic treatment of the sequencing data and taxonomic assignment of *Leishmania* spp. are available in the supplementary material.

ACKNOWLEDGMENTS

We would like to thank Cécile Cassan for her help in the laboratory work. *Leishmania* strains were provided by the Biological Ressource Center for *Leishmania* (Montpellier, France), the Parasitology and Mycology Laboratory of Cayenne Hospital (French Guiana), the Unit of Molecular Parasitology (Institute of Tropical Medicine, Antwerp, Belgium) and the Alexander von Humboldt Institute (Universidad Peruana Cayetano Heredia, Lima, Peru).

FINANCIAL SUPPORT

This work was supported by "Investissement d'Avenir" grants managed by Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; TULIP: ANR-10-LABX-41, ANR-11-IDEX-0002-02; IAEL: ANR-06-SEST-20) as well as the Inkermann fund hosted at the Fondation de France.

REFERENCES

Akhoundi, M., Downing, T., Votýpka, J., Kuhls, K., Lukeš, J., Cannet, A., Ravel, C., Marty, P., Delaunay, P., Kasbari, M., Granouillac, B., Gradoni, L. and Sereno, D. (in press). *Leishmania* infections: Molecular targets and diagnosis. *Molecular Aspects of Medicine*. doi:10.1016/j.mam.2016.11.012.

Akhoundi, M., Kuhls, K., Cannet, A., Votýpka, J., Marty, P., Delaunay, P. and Sereno, D. (2016). A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Neglected Tropical Diseases* 10, e0004349. doi:10.1371/journal.pntd.0004349.

Alvar, J., Vélez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J., Boer, M. and the WHO Leishmaniasis Control Team (2012). Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE* 7, e35671. doi:10.1371/journal.pone.0035671.

Angelici, M. C., Gramiccia, M. and Gradoni, L. (1989). Study on genetic polymorphism of *Leishmania infantum* through the analysis of restriction enzyme digestion patterns of kinetoplast DNA. *Parasitology* 99, 301–309.

Bastrenta, B., Buitrago, R., Vargas, F., Le Pont, F., Torrez, M., Flores, M., Mita, N. and Brenière, S. F. (2002). First evidence of transmission of *Leishmania* (*Viannia*) *lainsoni* in a Sub Andean region of Bolivia. *Acta Tropica* 83, 249–253. doi:10.1016/S0001-706X(02)00129-8.

Berzunza-Cruz, M., Rodríguez-Moreno, Á., Gutiérrez-Granados, G., González-Salazar, C., Stephens, C. R., Hidalgo-Mihart, M., Marina, C. F., Rebollar-Téllez, E. A., Bailón-Martínez, D., Balcells, C. D., Ibarra-Cerdeña, C. N., Sánchez-Cordero, V. and Becker, I. (2015). *Leishmania* (*L.*) *mexicana* infected bats in Mexico: novel potential reservoirs. *PLOS Neglected Tropical Diseases* 9, e0003438. doi:10.1371/journal.pntd.0003438.

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R. and Willerslev, E. (2007). The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* 2, e197. doi:10.1371/journal.pone.0000197.

Botilde, Y., Laurent, T., Quispe Tintaya, W., Chicharro, C., Cañavate, C., Cruz, I., Kuhls, K., Schönian, G. and Dujardin, J.-C. (2006). Comparison of molecular markers for strain typing of *Leishmania infantum*. *Infection, Genetics and Evolution* 6, 440–446. doi:10.1016/j.meegid.2006.02.003.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P. and Coissac, E. (2016). obitools: a unix-inspired software package for DNA metabarcoding. *Molecular ecology resources* 16, 176–182.

Brewster, S. and Barker, D. C. (2002). Analysis of minicircle classes in *Leishmania* (*Viannia*) species. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 96, S55–S63. doi:10.1016/S0035-9203(02)90052-0.

Brown, S. D., Collins, R. A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C. J. and Cruickshank, R. H. (2012). Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources* 12, 562–565.

Cássia-Pires, R., Boité, M. C., D'Andrea, P. S., Herrera, H. M., Cupolillo, E., Jansen, A. M. and Roque, A. L. R. (2014). Distinct *Leishmania* species infecting wild caviomorph rodents (Rodentia: Hystricognathi) from Brazil. *PLoS Neglected Tropical Diseases* 8, e3389. doi:10.1371/journal.pntd.0003389.

Ceccarelli, M., Galluzzi, L., Migliazzo, A. and Magnani, M. (2014). Detection and characterization of *Leishmania* (*Leishmania*) and *Leishmania* (*Viannia*) by SYBR Green-based real-time PCR and high resolution melt analysis targeting kinetoplast minicircle DNA. *PLOS ONE* 9, e88845. doi:10.1371/journal.pone.0088845.

Copeland, N. K. and Aronson, N. E. (2015). Leishmaniasis: treatment updates and clinical practice guidelines review. *Current Opinion in Infectious Diseases* 28, 426–437. doi:10.1097/QCO.0000000000000194.

de Almeida, M. E., Steurer, F. J., Koru, O., Herwaldt, B. L., Pieniazek, N. J. and da Silva, A. J. (2011). Identification of *Leishmania* spp. by molecular amplification and DNA sequencing analysis of a fragment of rRNA internal transcribed spacer 2. *Journal of Clinical Microbiology* 49, 3143–3149. doi:10.1128/JCM.01177-11.

de Oliveira Ramos Pereira, L. and Brandão, A. (2013). An analysis of trypanosomatids kDNA minicircle by absolute dinucleotide frequency. *Parasitology International* 62, 397–403. doi:10.1016/j.parint.2013.04.005.

Dufour, P., Pons, J.-M., Collinson, J. M., Gernigon, J., Dies, J. I., Sourrouille, P. and Crochet, P.-A. (2016). Multilocus barcoding confirms the occurrence of Elegant Terns in Western Europe. *Journal of Ornithology* 158, 351-361. doi: 10.1007/s10336-016-1380-0.

Dujardin, J.-C., Bañuls, A.-L., Llanos-Cuentas, A., Alvarez, E., DeDoncker, S., Jacquet, D., Le Ray, D., Arevalo, J. and Tibayrenc, M. (1995). Putative *Leishmania* hybrids in the Eastern Andean valley of Huanuco, Peru. *Acta Tropica* 59, 293–307. doi:10.1016/0001-706X(95)00094-U.

Esling, P., Lejzerowicz, F. and Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research* 43, 2513–2524. doi:10.1093/nar/gkv107.

Faye, B., Bañuls, A. L., Bucheton, B., Dione, M. M., Bassanganam, O., Hide, M., Dereure, J., Choisy, M., Ndiaye, J. L., Konaté, O., Claire, M., Senghor, M. W., Faye, M. N., Sy, I., Niang, A. A., Molez, J. F., Victoir, K., Marty, P., Delaunay, P., Knecht, R., Mellul, S., Diedhiou, S. and Gaye, O. (2010). Canine visceral leishmaniasis caused by *Leishmania infantum* in Senegal: risk of emergence in humans? *Microbes and Infection* 12, 1219–1225. doi:10.1016/j.micinf.2010.09.003.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., Percy, D. M., Hajibabaei, M. and Barrett, S. C. H. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLOS ONE* 3, e2802. doi:10.1371/journal.pone.0002802.

Gao, G., Kapushoc, S. T., Simpson, A. M., Thiemann, O. H. and Simpson, L. (2001). Guide RNAs of the recently isolated LEM125 strain of *Leishmania tarentolae*: an unexpected complexity. *RNA* 7, 1335–1347.

Gardener, P. J. and Howells, R. E. (1972). Isoenzyme variation in leishmanial parasites. *Journal of Protozoology* 19,.

Gibson, W., Crow, M. and Kearns, J. (1997). Kinetoplast DNA minicircles are inherited from both parents in genetic crosses of *Trypanosoma brucei*. *Parasitology Research* 83, 483–488. doi:10.1007/s004360050284.

Hajduk, S. and Ochsenreiter, T. (2010). RNA editing in kinetoplastids. *RNA Biology* 7, 229–236. doi:10.4161/rna.7.2.11393.

Harkins, K. M., Schwartz, R. S., Cartwright, R. A. and Stone, A. C. (2016). Phylogenomic reconstruction supports supercontinent origins for *Leishmania*. *Infection, Genetics and Evolution* 38, 101–109. doi:10.1016/j.meegid.2015.11.030.

Jensen, R. E. and Englund, P. T. (2012). Network news: The replication of kinetoplast DNA. *Annual Review of Microbiology* 66, 473–491. doi:10.1146/annurev-micro-092611-150057.

Koarashi, Y., Cáceres, A. G., Saca, F. M. Z., Flores, E. E. P., Trujillo, A. C., Alvares, J. L. A., Yoshimatsu, K., Arikawa, J., Katakura, K., Hashiguchi, Y. and Kato, H. (2016). Identification of causative *Leishmania* species in Giemsa-stained smears prepared from patients with cutaneous leishmaniasis in Peru using PCR-RFLP. *Acta Tropica* 158, 83–87. doi:10.1016/j.actatropica.2016.02.024.

Kocher, A., Thoisy, B. D., Catzeflis, F., Huguin, M., Valiere, S., Zinger, L., Bañuls, A.-L. and Murienne, J. (in press). Evaluation of short mitochondrial metabarcodes for the identification of Amazonian mammals. *Methods in Ecology and Evolution*. doi:10.1111/2041-210X.12729.

Kocher, A., Gantier, J.-C., Gaborit, P., Zinger, L., Holota, H., Valiere, S., Dusfour, I., Girod, R., Bañuls, A.-L. and Murienne, J. (2017). Vector soup: high-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources* 17, 172–182. doi:10.1111/1755-0998.12556.

Kreutzer, R. D., Semko, M. E., Hendricks, L. D. and Wright, N. (1983). Identification of *Leishmania* spp. by multiple isozyme analysis. *The American journal of tropical medicine and hygiene* 32, 703–715.

Lainson, R. and Shaw, J. J. (2010). New world leishmaniasis. *Topley and Wilson's microbiology and microbial infections*. doi:10.1002/9780470688618.taw0182.

Lee, S.-T., Tarn, C. and Chang, K.-P. (1993). Characterization of the switch of kinetoplast DNA minicircle dominance during development and reversion of drug resistance in *Leishmania*. *Molecular and Biochemical Parasitology* 58, 187–203. doi:10.1016/0166-6851(93)90041-U.

Marco, J. D., Uezato, H., Mimori, T., Barroso, P. A., Korenaga, M., Nonaka, S., Basombrío, M. A., Taranto, N. J. and Hashiguchi, Y. (2006). Are cytochrome B gene sequencing and polymorphism-specific polymerase chain reaction as reliable as multilocus enzyme electrophoresis for identifying *Leishmania* spp. from Argentina? *The American journal of tropical medicine and hygiene* 75, 256–260.

Marfurt, J., Nasereddin, A., Niederwieser, I., Jaffe, C. L., Beck, H.-P. and Felger, I. (2003). Identification and differentiation of *Leishmania* species in clinical samples by PCR amplification of the miniexon sequence and subsequent restriction fragment length polymorphism analysis. *Journal of Clinical Microbiology* 41, 3147–3153.

Mariette, J., Escudié, F., Allias, N., Salin, G., Noirot, C., Thomas, S. and Klopp, C. (2012). NG6: Integrated next generation sequencing storage and processing environment. *BMC genomics* 13, 462.

Maroli, M., Feliciangeli, M. D., Bichaud, L., Charrel, R. N. and Gradoni, L. (2013). Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Medical and Veterinary Entomology* 27, 123–147. doi:10.1111/j.1365-2915.2012.01034.x.

Noyes, H. A., Reyburn, H., Bailey, J. W. and Smith, D. (1998). A nested-PCR-based schizodeme method for identifying *Leishmania* kinetoplast minicircle classes directly from clinical samples and its application to the study of the epidemiology of *Leishmania tropica* in Pakistan. *Journal of clinical microbiology* 36, 2877–2881.

Paradis, E., Claude, J. and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics (Oxford, England)* 20, 289–290.

Pereira Júnior, A. M., Teles, C. B. G., de Azevedo dos Santos, A. P., de Souza Rodrigues, M., Marialva, E. F., Pessoa, F. A. C. and Medeiros, J. F. (2015). Ecological aspects and molecular detection of *Leishmania* DNA Ross (Kinetoplastida: Trypanosomatidae) in phlebotomine sandflies (Diptera: Psychodidae) in terra firme and várzea environments in the Middle Solimões Region, Amazonas State, Brazil. *Parasites & Vectors* 8,. doi:10.1186/s13071-015-0789-2.

R Core team (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012*. ISBN 3-900051-07-0.

Read, L. K., Lukeš, J. and Hashimi, H. (2015). Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley Interdisciplinary Reviews: RNA* n/a-n/a. doi:10.1002/wrna.1313.

Ready, P. D. (2013). Biology of phlebotomine sand flies as vectors of disease agents. *Annual Review of Entomology* 58, 227–250. doi:10.1146/annurev-ento-120811-153557.

Richini-Pereira, V. B., Marson, P. M., Hayasaka, E. Y., Victoria, C., da Silva, R. C. and Langoni, H. (2014). Molecular detection of *Leishmania* spp. in road-killed wild mammals in the Central Western area of the State of São Paulo, Brazil. *Journal of Venomous Animals and Toxins including Tropical Diseases* 20, 01–07.

Rioux, J. A., Lanotte, G., Serres, E., Pratlong, F., Bastien, P., Perieres, J. and others (1990). Taxonomy of *Leishmania*. Use of isoenzymes. Suggestions for a new classification. *Ann Parasitol Hum Comp* 65, 111–125.

Rodrigues, E. H. G., da Silva Soares, F. C., Werkhäuser, R. P., de Brito, M. E. F., Fernandes, O., Abath, F. G. C. and Brandão, A. (2013). The compositional landscape of minicircle sequences isolated from active lesions and scars of American cutaneous leishmaniasis. *Parasites & vectors* 6, 1.

Roque, A. L. R., Cupolillo, E., Marchevsky, R. S. and Jansen, A. M. (2010). *Thrichomys laurentius* (Rodentia; Echimyidae) as a Putative Reservoir of *Leishmania infantum* and *L. braziliensis*: Patterns of Experimental Infection. *PLOS Neglected Tropical Diseases* 4, e589. doi:10.1371/journal.pntd.0000589.

Rotureau, B. (2006). Ecology of the *Leishmania* species in the Guianan ecoregion complex. *The American journal of tropical medicine and hygiene* 74, 81–96.

Rougeron, V., De Meeûs, T. and Bañuls, A.-L. (in press). Reproduction in *Leishmania*: A focus on genetic exchange. *Infection, Genetics and Evolution*. doi:10.1016/j.meegid.2016.10.013.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* 4, 406–425.

Schnell, I. B., Bohmann, K. and Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources* 15, 1289–1303. doi:10.1111/1755-0998.12402.

Simon, S., Veron, V. and Carme, B. (2010). *Leishmania* spp. identification by polymerase chain reaction–restriction fragment length polymorphism analysis and its applications in French Guiana. *Diagnostic Microbiology and Infectious Disease* 66, 175–180. doi:10.1016/j.diagmicrobio.2009.08.013.

Simpson, L. (1987). The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. *Annual Review of Microbiology* 41, 363–380. doi:10.1146/annurev.mi.41.100187.002051.

Simpson, L. (1997). The genomic organization of guide RNA genes in kinetoplastid protozoa: several conundrums and their solutions. *Molecular and Biochemical Parasitology* 86, 133–141. doi:10.1016/S0166-6851(97)00037-6.

Telleria, J., Lafay, B., Virreira, M., Barnabé, C., Tibayrenc, M. and Svoboda, M. (2006). *Trypanosoma cruzi*: Sequence analysis of the variable region of kinetoplast minicircles. *Experimental Parasitology* 114, 279–288. doi:10.1016/j.exppara.2006.04.005.

Thomaz-Soccol, V., Lanotte, G., Rioux, J. A., Pratlong, F., Martini-Dumas, A. and Serres, E. (1993). Monophyletic origin of the genus *Leishmania* Ross, 1903. *Annales de parasitologie humaine et comparée* 68, 107–107.

Victoir, K., Bañuls, A. L., Arevalo, J., Llanos-Cuentas, A., Hamers, R., Noël, S., Doncker, S. D., Ray, D. L., Tibayrenc, M. and Dujardin, J. C. (1998). The gp63 gene locus, a target for genetic characterization of *Leishmania* belonging to subgenus *Viannia*. *Parasitology* 117, 1–13.

Weirather, J. L., Jeronimo, S. M., Gautam, S., Sundar, S., Kang, M., Kurtz, M. A., Haque, R., Schriefer, A., Talhari, S., Carvalho, E. M. and others (2011). Serial quantitative PCR assay for detection, species discrimination, and quantification of *Leishmania* spp. in human samples. *Journal of clinical microbiology* 49, 3892–3904.

Zelazny, A. M., Fedorko, D. P., Li, L., Neva, F. A. and Fischer, S. H. (2005). Evaluation of 7sl RNA gene sequences for the identification of *Leishmania* spp. *The American Journal of Tropical Medicine and Hygiene* 72, 415–420.

SUPPLEMENTARY MATERIAL

Appendix 1: PCR sensitivity assay

Ten strains, representative of six *Leishmania* species, were used to evaluate the sensitivity of our PCR assay (see Table below). Double-stranded DNA quantification was performed by fluorimetric analysis using the Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Each DNA sample was first diluted to 0.5 ng/µl and then serially ten-fold down to 0.05 fg/µl (to assess the detection capacity of the primers from 1 ng to 0.1 fg of template DNA).

Table: List of the strains used for the PCR sensitivity assay

Short code	WHO code	Species	Country
2007-24	MHOM/GF/2007/2007-24	*L. amazonensis*	Guyane
LV78	RAT/BR/72/LV78	*L. amazonensis*	Brésil
LH754	MHOM/PE/89/LH754	*L. braziliensis*	Pérou
M2904	MHOM/BR/75/M2904	*L. braziliensis*	Brésil
LC2288	MHOM/PE/91/LC2288	*L. lainsoni*	Pérou
TD2008	MHOM/GF/2008/TD2008	*L. guyanensis*	Guyane
ITMAP263	MHOM/MA/67/ITMAP263	*L. infantum*	Maroc
LEM417	MHOM/DZ/1982/LIPA59	*L. infantum*	Algérie
LEM2204	MDAS/BR/79/M5533	*L. naiffi*	Brésil
LEM5108	MHOM/GF/2005/LEM5108	*L. naiffi*	Guyane

A PCR assay was then conducted with the leishmini primers. Amplification was performed in 20 µL mixtures containing 2 µL of DNA template, 10 µL of AmpliTaq Gold PCR Master Mix® (5U.µL-1; Applied Biosystems, Foster City, CA, USA), 2.5 µL of each primer (5 µM), and nuclease-free water (Promega, Madison, WI, USA). The PCR mixture was denatured at 95°C (10 min) and followed by 35 cycles of 30s at 95 °C, 30s at 60°C and 15s at 72 °C, completed at 72 °C for 5 mins.

PCR product were subjected to electrophoresis in a 2% agarose gel containing SYBR Safe stain (Invitrogen), and visualized under UV light (see Figure below).

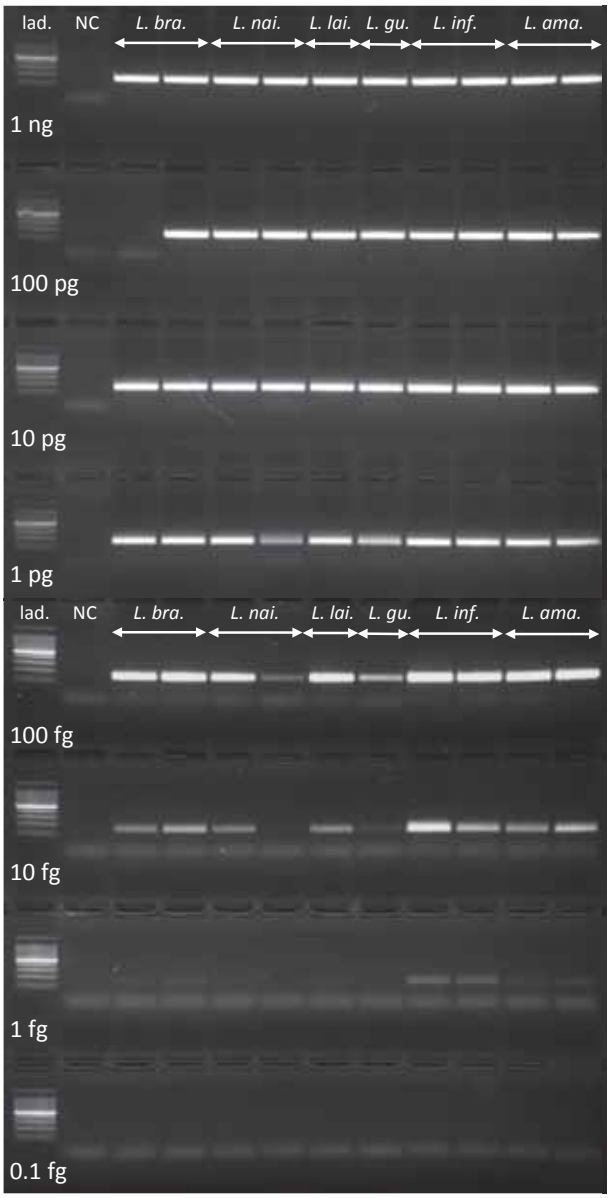


Figure: Agarose gel analysis of PCR products obtained with leishmini primers on a serial dilution of DNA from six Leishmania species

Appendix 2: Bash script used for the bioinformatic treatment of the sequencing data and taxonomic assignations with the OBITools package

```
# Alignment and merging of paired-end reads using illuminapairedend. The program
# assigns an alignment score to each resulting sequence based on the phred
# quality scores and the length of the aligned regions.
illuminapairedend --fasta-output -r readsR1.fastq readsR2.fastq \
> readsR1R2.fasta

# Reads assignment using ngsfilter. The program requires a table providing the
# information regarding the primer pair and the tag combination used for each
# sample (see the OBITools documentation for more details). This step will add
# an attribute to each sequence containing the name of the corresponding sample.
# Other information can be added.

ngsfilter -t ngsfilter.tab -e 2 --nuc readsR1R2.fasta > readsR1R2_ngsfilt.fasta

# Removal of low quality reads using obigrep (alignment scores<50, containing Ns
# or shorter than 50bp)

obigrep -s '^[acgt]+$' -l 50 -p 'score>=50' readsR1R2_ngsfilt.fasta \
> readsR1R2_ngsfilt_lowqual.fasta

# Dereplication of the sequences using obiuniq (regroups every identical reads
# assigned to the same sample into one sequence and keeps the coverage information)

obiuniq -c sample readsR1R2_ngsfilt_lowqual.fasta \
> readsR1R2_ngsfilt_lowqual_derep.fasta

# Check for the maximum sequence coverage found among sequence assigned to non-used
# tag combinations (i.e. resulting from tag-switching events), using obistat. These
# sequences can be selected with obigrep thanks to an attribute added in the
# ngsfilter tab. Here, we used the value 'NU' (not used) in the ID attribute

obigrep -a 'ID:NU' readsR1R2_ngsfilt_lowqual_derep.fasta | obistat -M count

# Remove sequences with a count lower than the maximum coverage found among
# sequence resulting from tag-switching events (as identified in the previous
# step, 1174 in our case)

obigrep -p 'count>1174' readsR1R2_ngsfilt_lowqual_derep.fasta \
> readsR1R2_ngsfilt_lowqual_derep_min1174.fasta

# Taxonomic assignations of each sequence using ecotag. The program requires a
# list of reference sequences in fasta format annotated with GenBank taxids
# (taxid=XX) and the corresponding genbank taxonomy dump (or ecoPCR taxonomy
database)

ecotag -R minicircles_refDB.fasta -d genbank_rXX \
readsR1R2_ngsfilt_lowqual_derep_min1174.fasta > minicircles_ecotag.fasta

# Output assignation statistics in a tab-delimited file
obistat -c sample -c species_name minicircles_ecotag.fasta \
> minicircles_assignation_stat.txt

# Run the small R script provided in the supplementary material to output an
identification table

Rscript Suppinfo2.R
```

Appendix 3: R script to create an identification table

```
#DATA IMPORT (obistat output)
#############
data=read.table("minicircles_assignation_stat.txt",header=T,sep="\t",colClasses=c("
character","character","numeric","numeric"))

#CREATION OF THE IDENTIFICATION TABLE
#############

#Create an empty dataframe
#==============================
#row names are the names of the sample
#column names depend on the list of species identified among the samples
sp_names=sort(unique(data[,2]))[sort(unique(data[,2]))!="None"]
minicircles_identification=as.data.frame(matrix(0,nrow=length(unique(data[,1])),nco
l=length(sp_names)+3))
rownames(minicircles_identification)=unique(data[,1])
colnames(minicircles_identification)=c("Identification","Score",sp_names,"None")

#Computing of the proportion of sequences assigned to each species in each sample
#==============================
#takes into account the abundance of each sequence
for (i in unique(data[,1])){
  for (j in unique(data[,2])){
    if (length(data$total[data[,1]==i&data[,2]==j])>0)

minicircles_identification[i,j]=data$total[data[,1]==i&data[,2]==j]/sum(data$total[
data[,1]==i])
  }
}

#Samples identification
#==============================
#the sample is identified as the species to wich the highest proportion of
sequences has been assigned
minicircles_identification$Identification=colnames(minicircles_identification)[max.
col(minicircles_identification)]

#Computing of an identification score
#==============================
#the identification score is defined as the proportion of the identified species
among all specific
#assignations (i.e. excluding sequences that did not find a specific match in the
references)
minicircles_identification$Score=apply(minicircles_identification[,sp_names],1,FUN=
max)/rowSums(minicircles_identification[sp_names])

#Rename columns and ouptut the table in a tab-delimited file
#==============================
colnames(minicircles_identification)=c("Identification","Score",paste("prop.",sp_na
mes),"None")
write.table(minicircles_identification,file =
"minicircles_identifications.txt",col.names = T,row.names = T,sep='\t',quote = F)
```

PRELIMINARY RESULTS:

TESTING THE DILUTION EFFECT HYPOTHESIS FOR ZOONOTIC CUTANEOUS LEISHMANIASES

INTRODUCTION

The "dilution effect" hypothesis suggests that biodiversity contributes to lower the transmission of infectious diseases through the presence of non-competent hosts in the ecosystems (Keesing *et al.* 2010). In the case of vector-borne diseases, for which the concept was initially developed, the dilution effect exerts when infected vectors feed on non-competent hosts, constituting a dead-end in the pathogen transmission cycle (Schmidt & Ostfeld 2001; LoGiudice *et al.* 2003). Through this mechanism, higher disease prevalence is expected in the context of low diversity if (i) vectors are relatively generalist and their abundance do not increase overly with host diversity and (ii) hosts that are more competent to transmit the given pathogen tend to persist in poorer communities (Ostfeld & Keesing 2012). It has been claimed that these conditions should be frequently met, leading to an overall negative effect of biodiversity on disease transmission (Civitello *et al.* 2015), but this remains a controversial question (Randolph & Dobson 2012; Salkeld *et al.* 2013; Wood *et al.* 2014). While modeling and experimental works have largely proved the theoretical validity of the dilution effect hypothesis (Dobson 2004; Rudolf & Antonovics 2005; Johnson *et al.* 2008, 2009), its demonstration in natural settings has been restricted to few emblematic systems, and often relied on indirect evidence. For animal vector-borne diseases, most studies concerned Lyme disease and West Nile fever (Allan *et al.* 2003; LoGiudice *et al.* 2003; Ezenwa *et al.* 2006; Swaddle & Calos 2008).

Leishmaniases are a set of diseases caused by protozoa belonging to the genus *Leishmania* Ross, 1903 (Kinetoplastida: Trypanosoatidae), and transmitted to humans and other mammals by phlebotomine sand flies (Diptera: Psychodidae). They are distributed worldwide and estimated to cause more than a million cases each year (Alvar *et al.* 2012). Leishmaniases are frequently considered to be re-emerging as a consequence of human-induced environmental changes such as urbanization, deforestation and natural habitat encroachment (Desjeux 2001, 2004; Maroli *et al.* 2013). However, the consequences of anthropogenic perturbations on their transmission cycles have never been assessed under the dilution effect hypothesis. In the New world, leishmaniases are caused by a variety of *Leishmania* species, most of which are characterized by a primary sylvatic cycle involving several but distinct sand fly vectors and wild mammal reservoirs (Lainson & Shaw 2010). Though, these diseases are likely to be impacted by modifications of ecological communities, and represent an interesting model for biodiversity-disease research with a multi-pathogen, multi-vector and multi-host perspective.

OBJECTIVES AND METHODS

In this preliminary work, we investigated the effect of human-induced perturbation on the sylvatic transmission cycle of leishmaniases in the region of Saint-Georges de l'Oyapock in French Guiana. More specifically, we explored the dilution effect hypothesis by evaluating (i) the abundance of sand fly vectors, (ii) the proportion of mammal reservoirs unveiled by sand fly blood meals and (iii) the prevalence of *Leishmania* parasites along the anthropogenic pressure gradient presented in the second article of chapter 3 (Figure 1). Site "SG1", the most impacted site, was located along the road leading to Brazil, close to the Oyapock river, at 1.2 km from Saint-Georges and exhibited an HFP index of 32. Sites "SG6" and "SG30" were located along the national road 2 (constructed 15 years ago) at 6.8 km and 30.3 km from Saint-Georges and presented an HFP index of 26 and 21 respectively (Figure 1). Dipteran blood meal analyses presented in the previous chapter unveiled a decline in vertebrate diversity along this gradient (Figure 2).

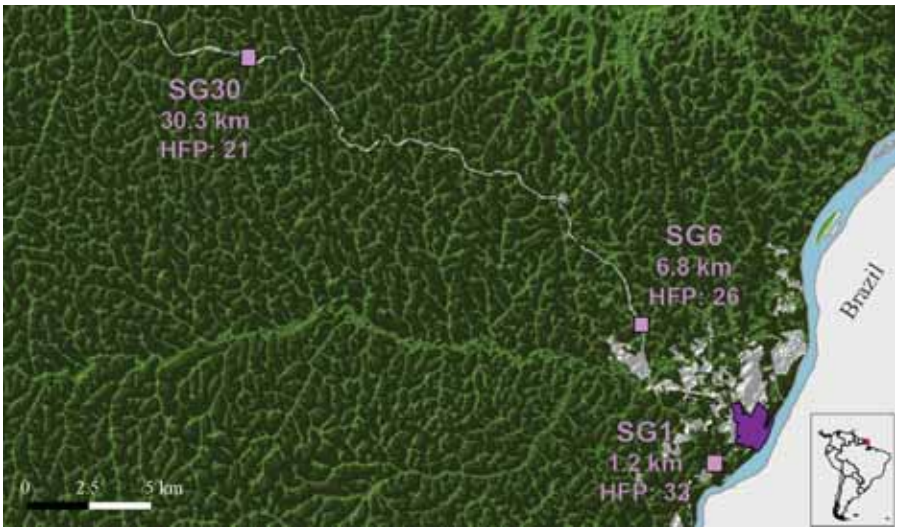


Figure 1: Geographical localisation of sampling sites in the area of Saint-Georges de l'Oyapock, French Guyana. Human footprint index values (HFP, de Thoisy et al. 2010) and distance from town are indicated. Forested areas are represented in green. Saint-Georges town is displayed in dark purple.

We used high-throughput sequencing of kDNA minicircle amplicons (cf. Article 1) to detect and identify *Leishmania* spp. in sand fly pools. Sand fly infection rates were then estimated based on pool results using a Bayesian method implemented in the package R 'prevalence'. The resulting values were weighted by mean sand fly abundances in each site to estimate the mean abundance of infected sand flies, as a proxy for disease risk. Characterization of sand fly communities by metabarcoding and sand fly blood meal analyses presented in the previous chapters are used to interpret the results in the light of current eco-epidemiological knowledge.

RESULTS AND DISCUSSION

Sand fly vector abundance

A total of 9.936 female sand flies were collected. After computing modified Z-scores and removing outlier traps, the remaining samples contained 8.039 specimens. The average number of specimens caught per trap-night increased significantly along the anthropogenic pressure gradient (41, 107 and 133 in SG30, SG6 and SG1 respectively). However, this pattern may not have direct epidemiological consequences since species that are not known to be vectors of *Leishmania* were found in very large numbers in the most perturbed site (see Chapter 2). As a more relevant indicator, the abundance of known or suspected vectors of *Leishmania* detected in the samples (*i.e. Nyssomyia umbratilis, Psychodopygus ayrozai, Psychodopygus s. maripaensis, Trichophoromyia ubiquitalis*; Rotureau 2006; Fouque *et al.* 2007) was estimated. The latter varied less between sites. In particular, the difference observed between SG1 and SG30 was very little and not significant (Figure 3A).

A dilution effect cannot be expected if host diversity overly promotes vector abundance. This was apparently not the case here even though estimations of species abundances based on metabarcoding have to be taken with precaution because of potential bias linked to differences in biomass and mitochondrial copy numbers between species, as well as variability in PCR primer binding sites (see Chapter 2, Article 1). The latter seems however to be null within Phlebotominae (based on the mitogenomic data provided in Chapter 1). Furthermore, the list of species used to estimate overall vector abundance should not be regarded as exhaustive (it has been more than ten years since the last study aiming at identifying potential vector species in French Guiana was conducted). For instance, species responsible for the transmission of *L. braziliensis* in the region remain unidentified so far.

Proportion of Leishmania reservoirs in sand fly blood meals

A central assumption of the dilution effect hypothesis is that species that are the most competent for pathogen transmission are also those that tend to persist when biodiversity is lost (Ostfeld & Keesing 2012). This supposed relationship is thought to be due to the existence of life history trade-off between resilience and competence or because of local adaptation of parasites to most common hosts (Joseph *et al.* 2013; Johnson *et al.* 2015). Although some studies do support the existence of such associations (Cardillo *et al.* 2008; Han *et al.* 2015), other patterns are sometime observed (Cooper *et al.* 2012; Young *et al.* 2013). In the case of American cutaneous leishmaniases, most known mammal reservoirs

such as armadillos, sloths, opossums and small rodents do indeed seem to be disturbance-tolerant species (Arias *et al.* 1981; Lopes & Ferrari 2000; Michalski & Peres 2007).

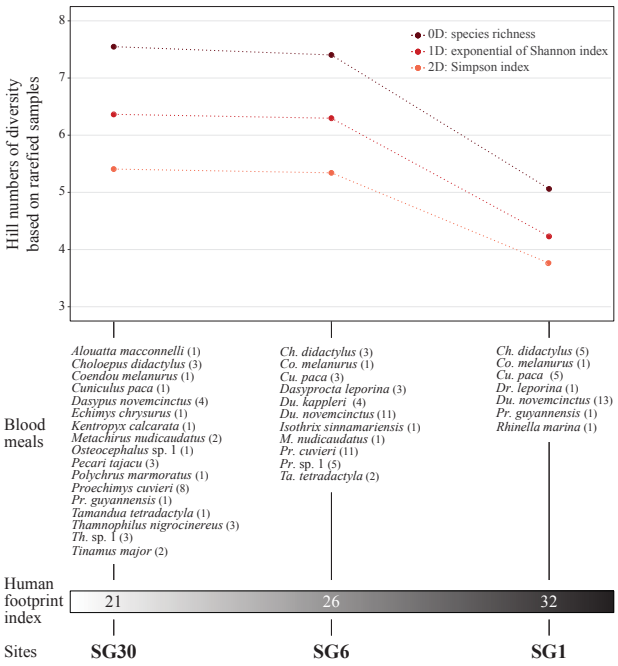


Figure 2: Vertebrates identified in dipteran blood meals (numbers in brackets indicate occurrences per trap-night) and estimations of Hill numbers of diversity (^{q}D) on rarefied blood meal pools among sites, along an anthropogenic pressure gradient defined by the human footprint index (de Thoisy *et al.* 2010).

Analyses of sand fly blood meals allowed identifying several known mammalian reservoirs of *Leishmania spp.* (Rotureau 2006): *Choloepus didactylus*, *Tamandua tetradactyla*, *Proechimys cuvieri*, *Proechimys guyannensis*, *Dasypus novemcinctus*, *Cuniculus paca*, and *Dasyprocta leporina*. Site comparisons indicated that the proportion of these hosts in sand fly blood meals increased along the perturbation gradient (62.5%, 83.3%% and 99.3% in SG30, SG6 and SG1 respectively; Figure 3B). These results seem to support the previously exposed assumption. However, it could be argued that common and resilient mammals have more chance to be caught and screened for *Leishmania* parasites than rarer ones. Therefore, the current list of suspected *Leishmania* reservoirs may be biased towards these species without any true relationship between resilience and competence. Further research is required in the light of these considerations to better define the major reservoirs locally.

Parasite prevalence

Leishmania spp. were detected in 18 out of 188 sand fly pools. Sand fly infection rate was estimated to be 0.23% (95% CI: 0.13-0.35%; Table 1), which was low compared to the most recent conducted in French Guiana (Fouque *et. al* 2007). Four *Leishmania* species were identified: *L. guyanensis*, *L. braziliensis*, *L. lainsoni* and *L. naiffi*. Surprisingly, *L. braziliensis* was the most frequent species. Although the latter has been reported as emerging

in French Guiana for several years (Simon *et al.* 2010; Martin-Blondel *et al.* 2015), *L. guyanensis* remains by far the most frequent agent of human leishmaniasis in this region (Simon *et al.* 2017). Nevertheless, sand fly infection rates do not directly reflect disease risk for humans. For instance, *L. lainsoni* and *L. naiffi* cases are rare, probably because of the low anthropophily of their sand fly vectors (Lainson & Shaw 2010).

Table 1: Sand fly infection rate (%) with Leishmania spp. based on detection on pools. Values obtained for each Leishmania species may not sum to the overall rate due to the estimation method.

	L. braziliensis	*L. guyanensis*	*L. naiffi*	*L. lainsoni*	**All**
SG30	0.5 (0.22-0.91)	0	0.25 (0.07-0.54)	0	0.65 (0.31-1.11)
SG6	0	0	0	0.17 (0.05-0.35)	0.17 (0.05-0.35)
SG1	0.12 (0.03-0.26)	0.21 (0.08-0.4)	0	0	0.28 (0.12-0.49)
All	0.11 (0.05-0.2)	0.07 (0.02-0.14)	0.05 (0.01-0.12)	0.07 (0.02-0.14)	0.23 (0.13-0.35)

Computation of *Leishmania* spp. prevalence by site indicated infection rates of 0.65% in SG30 (95% CI: 0.31-1.11%), 0.17% in SG6 (95% CI: 0.05-0.35%) and 0.28% in SG1 (95% CI: 0.12-0.49%). However, as emphasized by Randolph & Dobson (2012), the abundance (rather than rate) of infected sand flies represents a better proxy for disease risk. Estimation of the latter showed a different trend, due to difference of overall sand fly abundance between sites, with a maximum in SG1 (Figure 3C). This result is, together with estimations of sand fly vector abundance and proportion of mammal reservoirs in sand fly blood meals, seems to support the dilution effect hypothesis.

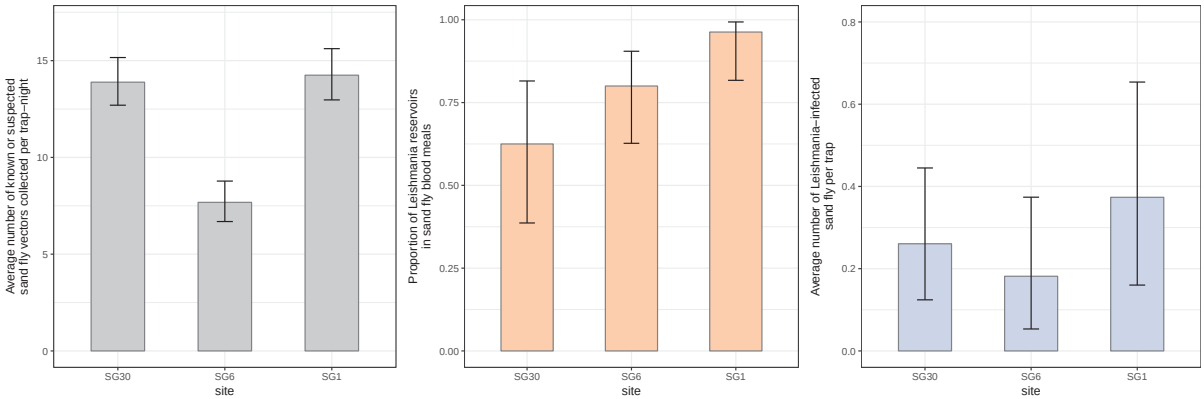


Figure 3: (A) Mean abundance of known sand fly vectors estimated with metabarcoding, (B) Proportion of known mammal reservoirs identified in sand fly blood meals, (C) Average number of infected sand fly per trap-night estimated with detection on pools

Nevertheless, the very low levels of prevalence observed hampered precise estimations, despite relatively high sampling efforts. Indeed, notwithstanding our findings, the presented

data do not allow to affirm significant difference in sand fly infection rates between sites (as showed by the wide overlaps of confidence intervals). The implementation of such study in areas or periods undergoing higher *Leishmania* transmission would be helpful in this regard. In any case, the inclusion of more sampling localities would be required to test the generality of the phenomenon. More detailed and mechanistic analyses would also be necessary. In particular, it would be relevant to consider each *Leishmania* species separately since these may have really distinct ecology. However, this cannot be done without more extensive data and a better knowledge of basic eco-epidemiological features such as vector-host associations and reservoir competence.

REFERENCES

Akhoundi M, Kuhls K, Cannet A *et al.* (2016) A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Neglected Tropical Diseases*, **10**, e0004349.

Allan BF, Keesing F, Ostfeld RS (2003) Effect of forest fragmentation on Lyme disease risk. *Conservation Biology*, **17**, 267–272.

de Almeida ME, Steurer FJ, Koru O *et al.* (2011) Identification of *Leishmania* spp. by molecular amplification and DNA sequencing analysis of a fragment of rRNA internal transcribed spacer 2. *Journal of Clinical Microbiology*, **49**, 3143–3149.

Alvar J, Vélez ID, Bern C *et al.* (2012) Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE*, **7**, e35671.

Angelici MC, Gramiccia M, Gradoni L (1989) Study on genetic polymorphism of *Leishmania infantum* through the analysis of restriction enzyme digestion patterns of kinetoplast DNA. *Parasitology*, **99**, 301–309.

Arias JR, Naiff RD, Miles MA, de Souza AA (1981) The opossum, Didelphis marsupialis (Marsupialia: Didelphidae), as a reservoir host of Leishmania braziliensis guyanensis in the Amazon Basin of Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **75**, 537–541.

Bañuls A-L, Dujardin J-C, Guerrini F *et al.* (2000) Is *Leishmania* (*Viannia*) *peruviana* a distinct species? A MLEE/RAPD evolutionary genetics answer. *Journal of Eukaryotic Microbiology*, **47**, 197–207.

Bañuls A-L, Hide M, Prugnolle F (2007) *Leishmania* and the leishmaniases: A parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. *Advances in Parasitology*, **64**, 1–109.

Bastrenta B, Buitrago R, Vargas F *et al.* (2002) First evidence of transmission of *Leishmania* (*Viannia*) *lainsoni* in a Sub Andean region of Bolivia. *Acta Tropica*, **83**, 249–253.

Berzunza-Cruz M, Rodríguez-Moreno Á, Gutiérrez-Granados G *et al.* (2015) *Leishmania* (*L.*) *mexicana* infected bats in Mexico: novel potential reservoirs (PA Bates, Ed,). *PLOS Neglected Tropical Diseases*, **9**, e0003438.

Binladen J, Gilbert MTP, Bollback JP *et al.* (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE*, **2**, e197.

Botilde Y, Laurent T, Quispe Tintaya W *et al.* (2006) Comparison of molecular markers for strain typing of *Leishmania infantum*. *Infection, Genetics and Evolution*, **6**, 440–446.

Boyer F, Mercier C, Bonin A *et al.* (2016) obitools: a unix-inspired software package for DNA metabarcoding. *Molecular ecology resources*, **16**, 176–182.

Brewster S, Barker DC (2002) Analysis of minicircle classes in *Leishmania* (*Viannia*) species. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **96**, S55–S63.

Brown SD, Collins RA, Boyer S *et al.* (2012) Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, **12**, 562–565.

Camacho C, Coulouris G, Avagyan V *et al.* (2009) BLAST+: architecture and applications. *BMC bioinformatics*, **10**, 421.

Cardillo M, Mace GM, Gittleman JL *et al.* (2008) The predictability of extinction: biological and external correlates of decline in mammals. *Proceedings of the Royal Society of London B: Biological Sciences*, **275**, 1441–1448.

Cássia-Pires R, Boité MC, D'Andrea PS *et al.* (2014) Distinct *Leishmania* species infecting wild caviomorph rodents (Rodentia: Hystricognathi) from Brazil (G Schönian, Ed,). *PLoS Neglected Tropical Diseases*, **8**, e3389.

Ceccarelli M, Galluzzi L, Migliazzo A, Magnani M (2014) Detection and characterization of *Leishmania* (*Leishmania*) and *Leishmania* (*Viannia*) by SYBR Green-based real-time PCR and high resolution melt analysis targeting kinetoplast minicircle DNA. *PLOS ONE*, **9**, e88845.

Civitello DJ, Cohen J, Fatima H *et al.* (2015) Biodiversity inhibits parasites: Broad evidence for the dilution effect. *Proceedings of the National Academy of Sciences*, 201506279.

Cooper N, Kamilar JM, Nunn CL (2012) Host Longevity and Parasite Species Richness in Mammals. *PLOS ONE*, **7**, e42190.

Copeland NK, Aronson NE (2015) Leishmaniasis: treatment updates and clinical practice guidelines review. *Current Opinion in Infectious Diseases*, **28**, 426–437.

Desjeux P (2001) The increase in risk factors for leishmaniasis worldwide. *Transactions of The Royal Society of Tropical Medicine and Hygiene*, **95**, 239–243.

Desjeux P (2004) Leishmaniasis: current situation and new perspectives. *Comparative Immunology, Microbiology and Infectious Diseases*, **27**, 305–318.

Dobson A (2004) Population Dynamics of Pathogens with Multiple Host Species. *The American Naturalist*, **164**, S64–S78.

Dufour P, Pons J-M, Collinson JM *et al.* (2016) Multilocus barcoding confirms the occurrence of Elegant Terns in Western Europe. *Journal of Ornithology*, 1–11.

Dujardin J-C, Bañuls A-L, Llanos-Cuentas A *et al.* (1995) Putative *Leishmania* hybrids in the Eastern Andean valley of Huanuco, Peru. *Acta Tropica*, **59**, 293–307.

Esling P, Lejzerowicz F, Pawlowski J (2015) Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, **43**, 2513–2524.

Ezenwa VO, Godsey MS, King RJ, Guptill SC (2006) Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 109–117.

Faye B, Bañuls AL, Bucheton B *et al.* (2010) Canine visceral leishmaniasis caused by *Leishmania infantum* in Senegal: risk of emergence in humans? *Microbes and Infection*, **12**, 1219–1225.

Fazekas AJ, Burgess KS, Kesanakurti PR *et al.* (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLOS ONE*, **3**, e2802.

Fouque F, Gaborit P, Issaly J *et al.* (2007) Phlebotomine sand flies (Diptera: Psychodidae) associated with changing patterns in the transmission of the human cutaneous leishmaniasis in French Guiana. *Memórias do Instituto Oswaldo Cruz*, **102**, 35–40.

Fraga J, Montalvo AM, De Doncker S, Dujardin J-C, Van der Auwera G (2010) Phylogeny of *Leishmania* species based on the heat-shock protein 70 gene. *Infection, Genetics and Evolution*, **10**, 238–245.

Fraga J, Montalvo AM, Van der Auwera G *et al.* (2013) Evolution and species discrimination according to the *Leishmania* heat-shock protein 20 gene. *Infection, Genetics and Evolution*, **18**, 229–237.

Gao G, Kapushoc ST, Simpson AM, Thiemann OH, Simpson L (2001) Guide RNAs of the recently isolated LEM125 strain of *Leishmania tarentolae*: an unexpected complexity. *RNA*, **7**, 1335–1347.

Gardener PJ, Howells RE (1972) Isoenzyme variation in leishmanial parasites. *Journal of Protozoology*, **19**.

Gibson W, Crow M, Kearns J (1997) Kinetoplast DNA minicircles are inherited from both parents in genetic crosses of *Trypanosoma brucei*. *Parasitology Research*, **83**, 483–488.

Graça GC da, Volpini AC, Romero GAS *et al.* (2012) Development and validation of PCR-based assays for diagnosis of American cutaneous leishmaniasis and identification of the parasite species. *Memórias do Instituto Oswaldo Cruz*, **107**, 664–674.

Greif G, Rodriguez M, Reyna-Bello A, Robello C, Alvarez-Valin F (2015) Kinetoplast adaptations in American strains from *Trypanosoma vivax*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **773**, 69–82.

Hajduk S, Ochsenreiter T (2010) RNA editing in kinetoplastids. *RNA Biology*, **7**, 229–236.

Han BA, Schmidt JP, Bowden SE, Drake JM (2015) Rodent reservoirs of future zoonotic diseases. *Proceedings of the National Academy of Sciences*, **112**, 7039–7044.

Harkins KM, Schwartz RS, Cartwright RA, Stone AC (2016) Phylogenomic reconstruction supports supercontinent origins for *Leishmania*. *Infection, Genetics and Evolution*, **38**, 101–109.

Jensen RE, Englund PT (2012) Network news: The replication of kinetoplast DNA. *Annual Review of Microbiology*, **66**, 473–491.

Johnson PTJ, Hartson RB, Larson DJ, Sutherland DR (2008) Diversity and disease: community structure drives parasite transmission and host fitness. *Ecology Letters*, **11**, 1017–1026.

Johnson PT., Lund PJ, Hartson RB, Yoshino TP (2009) Community diversity reduces *Schistosoma mansoni* transmission, host pathology and human infection risk. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1657–1663.

Johnson PTJ, Ostfeld RS, Keesing F (2015) Frontiers in research on biodiversity and disease. *Ecology Letters*, **18**, 1119–1133.

Joseph MB, Mihaljevic JR, Orlofske SA, Paull SH (2013) Does life history mediate changing disease risk when communities disassemble? *Ecology Letters*, **16**, 1405–1412.

Kearse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Keesing F, Belden LK, Daszak P *et al.* (2010) Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, **468**, 647–652.

Koarashi Y, Cáceres AG, Saca FMZ *et al.* (2016) Identification of causative *Leishmania* species in Giemsa-stained smears prepared from patients with cutaneous leishmaniasis in Peru using PCR-RFLP. *Acta Tropica*, **158**, 83–87.

Kocher A, Gantier J-C, Gaborit P *et al.* (2017) Vector soup: high-throughput identification of Neotropical phlebotomine sand flies using metabarcoding. *Molecular Ecology Resources*, **17**, 172–182.

Kocher A, Gantier J-C, Holota H *et al.* (2016) Complete mitochondrial genome of *Lutzomyia* (*Nyssomyia*) *umbratilis* (Diptera: Psychodidae), the main vector of *Leishmania guyanensis*. *Mitochondrial DNA*, **27**, 4219–4221.

Kocher A, Guilbert É, Lhuillier É, Murienne J (2015) Sequencing of the mitochondrial genome of the avocado lace bug *Pseudacysta perseae* (Heteroptera, Tingidae) using a genome skimming approach. *Comptes Rendus Biologies*, **338**, 149–160.

Kocher A, Kamilari M, Lhuillier E *et al.* (2014) Shotgun assembly of the assassin bug *Brontostoma colossus* mitochondrial genome (Heteroptera, Reduviidae). *Gene*, **552**, 184–194.

Kocher A, Thoisy BD, Catzeflis F *et al.* (in press) Evaluation of short mitochondrial metabarcodes for the identification of Amazonian mammals. *Methods in Ecology and Evolution*.

Kreutzer RD, Semko ME, Hendricks LD, Wright N (1983) Identification of *Leishmania* spp. by multiple isozyme analysis. *The American journal of tropical medicine and hygiene*, **32**, 703–715.

Lainson R, Shaw JJ (2010) New world leishmaniasis. *Topley and Wilson's microbiology and microbial infections*.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Lee S-T, Tarn C, Chang K-P (1993) Characterization of the switch of kinetoplast DNA minicircle dominance during development and reversion of drug resistance in *Leishmania*. *Molecular and Biochemical Parasitology*, **58**, 187–203.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

LoGiudice K, Ostfeld RS, Schmidt KA, Keesing F (2003) The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. *Proceedings of the National Academy of Sciences*, **100**, 567–571.

Lopes MA, Ferrari SF (2000) Effects of human colonization on the abundance and diversity of mammals in eastern Brazilian Amazonia. *Conservation Biology*, **14**, 1658–1665.

Marco JD, Uezato H, Mimori T *et al.* (2006) Are cytochrome B gene sequencing and polymorphism-specific polymerase chain reaction as reliable as multilocus enzyme electrophoresis for identifying *Leishmania* spp. from Argentina? *The American journal of tropical medicine and hygiene*, **75**, 256–260.

Marfurt J, Nasereddin A, Niederwieser I *et al.* (2003) Identification and differentiation of *Leishmania* species in clinical samples by PCR amplification of the miniexon sequence and subsequent restriction fragment length polymorphism analysis. *Journal of Clinical Microbiology*, **41**, 3147–3153.

Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L (2013) Phlebotomine sandflies and the spreading of leishmaniases and other diseases of public health concern. *Medical and Veterinary Entomology*, **27**, 123–147.

Martin-Blondel G, Iriart X, El Baidouri F *et al.* (2015) Outbreak of *Leishmania braziliensis* Cutaneous Leishmaniasis, Saül, French Guiana. *Emerging Infectious Diseases*, **21**, 892–894.

Michalski F, Peres CA (2007) Disturbance-Mediated Mammal Persistence and Abundance-Area Relationships in Amazonian Forest Fragments. *Conservation Biology*, **21**, 1626-1640

Noyes HA, Reyburn H, Bailey JW, Smith D (1998) A nested-PCR-based schizodeme method for identifying *Leishmania* kinetoplast minicircle classes directly from clinical samples and its application to the study of the epidemiology of *Leishmania tropica* in Pakistan. *Journal of clinical microbiology*, **36**, 2877–2881.

de Oliveira Ramos Pereira L, Brandão A (2013) An analysis of trypanosomatids kDNA minicircle by absolute dinucleotide frequency. *Parasitology International*, **62**, 397–403.

Ostfeld RS, Keesing F (2012) Effects of Host Diversity on Infectious Disease. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 157–182.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics (Oxford, England)*, **20**, 289–290.

Pereira Júnior AM, Teles CBG, de Azevedo dos Santos AP *et al.* (2015) Ecological aspects and molecular detection of *Leishmania* DNA Ross (Kinetoplastida: Trypanosomatidae) in phlebotomine sandflies (Diptera: Psychodidae) in terra firme and várzea environments in the Middle Solimões Region, Amazonas State, Brazil. *Parasites & Vectors*, **8**.

R Core team (2014) *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012*. ISBN 3-900051-07-0.

Randolph SE, Dobson ADM (2012) Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology*, **139**, 847–863.

Read LK, Lukeš J, Hashimi H (2015) Trypanosome RNA editing: the complexity of getting U in and taking U out. *Wiley Interdisciplinary Reviews: RNA*, n/a-n/a.

Ready PD (2013) Biology of phlebotomine sand flies as vectors of disease agents. *Annual Review of Entomology*, **58**, 227–250.

Richini-Pereira VB, Marson PM, Hayasaka EY *et al.* (2014) Molecular detection of *Leishmania* spp. in road-killed wild mammals in the Central Western area of the State of São Paulo, Brazil. *Journal of Venomous Animals and Toxins including Tropical Diseases*, **20**, 01–07.

Rioux JA, Lanotte G, Serres E *et al.* (1990) Taxonomy of *Leishmania*. Use of isoenzymes. Suggestions for a new classification. *Ann Parasitol Hum Comp*, **65**, 111–125.

Rodrigues EHG, da Silva Soares FC, Werkhäuser RP *et al.* (2013) The compositional landscape of minicircle sequences isolated from active lesions and scars of American cutaneous leishmaniasis. *Parasites & vectors*, **6**, 1.

Rotureau B (2006) Ecology of the *Leishmania* species in the Guianan ecoregion complex. *The American journal of tropical medicine and hygiene*, **74**, 81–96.

Rougeron V, De Meeûs T, Bañuls A-L (in press) Reproduction in *Leishmania*: A focus on genetic exchange. *Infection, Genetics and Evolution*.

Rudolf VHW, Antonovics J (2005) Species Coexistence and Pathogens with Frequency- Dependent Transmission. *The American Naturalist*, **166**, 112–118.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, **4**, 406–425.

Salkeld DJ, Padgett KA, Jones JH (2013) A meta-analysis suggesting that the relationship between biodiversity and risk of zoonotic pathogen transmission is idiosyncratic (K Lafferty, Ed,). *Ecology Letters*, **16**, 679–686.

Schmidt KA, Ostfeld RS (2001) Biodiversity and the dilution effect in disease ecology. *Ecology*, **82**, 609–619.

Schnell IB, Bohmann K, Gilbert MTP (2015) Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, **15**, 1289–1303.

Schönian G, Mauricio I, Cupolillo E (2010) Is it time to revise the nomenclature of *Leishmania* ? *Trends in Parasitology*, **26**, 466–469.

Simon S, Nacher M, Carme B *et al.* (2017) Cutaneous leishmaniasis in French Guiana: revising epidemiology with PCR-RFLP. *Tropical Medicine and Health*, **45**, 5.

Simon S, Veron V, Carme B (2010) *Leishmania* spp. identification by polymerase chain reaction–restriction fragment length polymorphism analysis and its applications in French Guiana. *Diagnostic Microbiology and Infectious Disease*, **66**, 175–180.

Simpson L (1987) The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication, and evolution. *Annual Review of Microbiology*, **41**, 363–380.

Simpson L (1997) The genomic organization of guide RNA genes in kinetoplastid protozoa: several conundrums and their solutions. *Molecular and Biochemical Parasitology*, **86**, 133–141.

Swaddle JP, Calos SE (2008) Increased Avian Diversity Is Associated with Lower Incidence of Human West Nile Infection: Observation of the Dilution Effect (A Buckling, Ed,). *PLoS ONE*, **3**, e2488.

Telleria J, Lafay B, Virreira M *et al.* (2006) *Trypanosoma cruzi*: Sequence analysis of the variable region of kinetoplast minicircles. *Experimental Parasitology*, **114**, 279–288.

de Thoisy B, Richard-Hansen C, Goguillon B *et al.* (2010) Rapid evaluation of threats to biodiversity: human footprint score and large vertebrate species responses in French Guiana. *Biodiversity and Conservation*, **19**, 1567–1584.

Thomaz-Soccol V, Lanotte G, Rioux JA *et al.* (1993) Monophyletic origin of the genus *Leishmania* Ross, 1903. *Annales de parasitologie humaine et comparée*, **68**, 107–107.

Victoir K, Bañuls AL, Arevalo J *et al.* (1998) The gp63 gene locus, a target for genetic characterization of *Leishmania* belonging to subgenus *Viannia*. *Parasitology*, **117**, 1–13.

Wood CL, Lafferty KD, DeLeo G *et al.* (2014) Does biodiversity protect humans against infectious disease? *Ecology*, **95**, 817–832.

Young H, Griffin RH, Wood CL, Nunn CL (2013) Does habitat disturbance increase infectious disease risk for primates? (K Lafferty, Ed,). *Ecology Letters*, **16**, 656–663.

Zelazny AM, Fedorko DP, Li L, Neva FA, Fischer SH (2005) Evaluation of 7sl RNA gene sequences for the identification of *Leishmania* spp. *The American Journal of Tropical Medicine and Hygiene*, **72**, 415–420.

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research*, **18**, 821–829.

DISCUSSION GÉNÉRALE

L'objectif de cette thèse était d'étudier l'impact de la perturbation des habitats sur le cycle de transmission des leishmanioses en Guyane française. En particulier, nous avons exploré l'hypothèse de l'effet de dilution. Il s'agissait à la fois d'apporter une contribution au débat théorique concernant les liens existant entre biodiversité et maladies infectieuses, et de générer des connaissances d'intérêt sanitaire plus direct pour la région d'étude. En outre, notre travail à permis d'apporter une contribution à la connaissance de l'histoire évolutive des phlébotomes par la génération et l'analyse des premières données mitogénomiques existantes pour ce groupe. Les résultats de cette étude, présentés dans le premier chapitre, apportent des éléments intéressants pour le reste de la thèse bien qu'ils ne se rapportent pas à directement à sa question principale. En particulier, ils confortent globalement la classification de Galati qui sera donc employée préférentiellement dans les articles écrits par la suite.

Levée du verrou méthodologique

Le cœur de cette thèse est construit en deux étapes avec, dans un premier temps, le développement d'outils méthodologiques basés sur l'utilisation des technologies de séquençage haut-débit, pour la caractérisation des trois composantes du modèle épidémiologique : parasites, vecteurs et hôtes. La réalisation d'études empiriques est bien souvent limitée par des difficultés pratiques à l'acquisition de données. Ceci explique probablement le fait que peu de travaux aient été menés pour tester l'hypothèse de l'effet de dilution de manière directe et en conditions naturelles. Avec les techniques classiques, la mise à disposition de moyens humains et financiers considérables seraient nécessaires pour caractériser les communautés d'hôtes vertébrés et de phlébotomes vecteurs ainsi que les taux de prévalence parasitaire sur un grand nombre de sites et dans un temps raisonnable. Cette première phase de développement méthodologique constituait donc une étape nécessaire à la suite de l'étude.

Nous avons montré que le *metabarcoding* permet une caractérisation rapide des communautés de phlébotomes, ce qui nécessitait jusqu'alors des connaissances taxonomiques pointues et/ou un travail de laboratoire important. Ce travail ouvre de nombreuses perspectives pour l'étude écologique des phlébotomes. Il sera toutefois nécessaire d'étendre la base de référence moléculaire à d'avantage d'espèces pour permettre son utilisation dans d'autres régions. Dans le futur, l'adoption d'approches de type metagénomique ne nécessitant pas d'amplification PCR préalable devrait permettre de réduire les biais inhérents à cette

étape et qui limitent l'utilisation du metabarcoding pour des mesures quantitatives précises. L'emploi du *genome skimming* et des données génétiques générées dans le Chapitre 1 pourront être mis à profit à cet égard.

Nous avons ensuite étudié la fiabilité d'un marqueur mitochondrial court situé dans la petite sous-unité ribosomale (12S) pour l'analyse des repas sanguins de phlébotomes. La constitution d'une base de référence quasiment exhaustive pour les mammifères de Guyane a permis de démontrer la bonne résolution du marqueur, qui fournit des identifications au niveau spécifique dans la plupart des cas. Nos résultats confirment également l'utilité de l'ADN ribosomal pour la conception d'amorces disposant d'une grande couverture taxonomique en comparaison à des gènes codants tels que le COI, qui constitue le *barcode* standard pour les animaux. L'emploi de la méthode sur des phlébotomes et moustiques gorgés de sang récoltés dans des sites forestiers en Guyane à permis l'identification de diverses espèces de vertébrés incluant des mammifères arboricoles de taille variée, des oiseaux, des lézards et des amphibiens. Nos résultats permettent d'envisager l'analyse des repas sanguins de diptères comme une alternative intéressante pour caractériser les communautés d'hôtes sur les sites d'étude, ce qui nécessite traditionnellement la mise en place de méthodes observationnelles très contraignantes. Un autre avantage de la technique est qu'elle permet simultanément d'obtenir de précieuses informations sur les préférences trophiques des vecteurs de leishmanioses. Toutefois, la faible proportion de spécimens gorgés collectés, et les difficultés analytiques liées à la diversité des préférences trophiques des diptères représentent des limites qui nécessitent le développement de protocoles d'échantillonnage et d'outils statistiques spécialement adaptés.

Le troisième volet méthodologique de la thèse est constitué du développement d'une technique de détection et d'identification des *Leishmania* basée sur l'amplification et le séquençage haut débit des minicercles d'ADN kinétoplastiques. Nous montrons que la méthode permet de combiner une grande capacité de détection des parasites et la réalisation d'identifications spécifiques. L'approche n'est pas encore adaptée pour une utilisation routinière (dans un cadre clinique notamment), mais peut déjà être employée pour des études écologiques nécessitant l'analyse d'un grand nombre d'échantillons. Au regard de l'évolution rapide et de l'accessibilité grandissante des technologies de séquençage haut-débit, ce travail pourrait constituer un premier pas vers le développement d'une nouvelle génération d'outils moléculaires pour l'étude des *Leishmania*.

Effet des perturbations anthropiques sur le cycle de transmission des leishmanioses : une étude exploratoire

Dans un second temps, nous avons appliqué les outils développés précédemment à des échantillons récoltés sur le terrain afin d'aborder notre question de thèse. Nos résultats, présentés le long des trois derniers chapitres, permettent d'explorer différents aspects du système d'étude en relation avec la perturbation de la faune de vertébrés par l'activité humaine.

La caractérisation des communautés de phlébotomes par *metabarcoding* dans des sites forestiers ou la faune a été plus ou moins impactée par l'homme a permis l'identification de patrons congruents dans deux localités distantes. Sur les sites les plus perturbés, la présence d'une ou deux espèces largement dominantes résulte en une augmentation de l'abondance totale des phlébotomes, et en une homogénéisation de la communauté. De façon surprenante, quasiment aucun spécimen n'a été observé gorgé de sang parmi ces espèces, conduisant à une baisse significative de la proportion d'individus gorgés sur les sites les plus impactés. Ceci pourrait indiquer une faculté autogénique chez espèces (*i.e.* capacité à se reproduire sans la prise de repas sanguin), expliquant leur adaptabilité des milieux ou les hôtes vertébrés sont plus rares. Toutefois, cette hypothèse permet difficilement d'expliquer l'augmentation de l'abondance totale des phlébotomes sur les sites perturbés. L'existence potentielle d'autres effets indirects des modifications de la faune de vertébrés sur l'habitat et les ressources alimentaires végétales des phlébotomes complique l'interprétation mécanistique des patrons observés. Par ailleurs, il peut être difficile de découpler l'effet de la « défaunation » de celui d'autres types de perturbations, comme l'exploitation forestière, qui tendent à survenir dans les mêmes zones. Enfin, il ne peut pas être exclus que des émergences massives de ces espèces aient eu lieu par hasard sur les sites les plus perturbés lors de la période de collecte. Une extension de l'échantillonnage serait nécessaire pour confirmer la généralité du phénomène et étudier ses conséquences épidémiologiques le cas échéant. En outre, une caractérisation plus fine des habitats sur les lieux d'études pourrait bénéficier à la compréhension des mécanismes sous-jacents.

L'une des conditions centrales à la manifestation d'un effet de dilution est que les hôtes les plus compétents pour la transmission des pathogènes aient tendance à persister dans les communautés les plus pauvres en espèces. L'existence d'un tel patron pourrait s'expliquer par une adaptation locale des parasites aux hôtes les plus communs ou par l'existence de corrélations négatives entre les traits d'histoire vie liés à la résilience et les capacités

immunitaires des espèces. Dans le cas des leishmanioses, la plupart des mammifères réservoirs connus semblent effectivement tolérants aux perturbations, bien que certains soient strictement dépendants du milieu forestier. L'analyse des repas sanguins de phlébotomes à permis d'observer une augmentation de la proportion des espèces réservoirs le long du gradient d'anthropisation, conformément à l'hypothèse formulée. Cependant, tous les hôtes potentiels de *Leishmania* n'ont vraisemblablement pas été identifiés. Il faut en particulier considérer que les espèces résilientes aux perturbations anthropiques sont aussi généralement plus abondantes et répandues, et évidemment plus susceptible d'être observées proche de l'habitat humain. Ceci peut conduire à la détection de parasites et donc à la suspicion du rôle de réservoir de façon plus probable chez ces espèces en dépit de tout lien existant entre résilience et compétence. D'avantage d'études visant à l'identification non biaisée des hôtes réservoirs de leishmanioses sont nécessaires à la compréhension précise des cycles de transmission.

La détection des *Leishmania* dans les pools de phlébotomes a permis l'estimation de taux de prévalence sur chaque site. Les résultats indiquent que l'abondance des vecteurs infectés, pouvant être employée comme une mesure du risque infectieux, était plus importante sur le site le plus perturbé. Ces résultats, semblent donc globalement conforter l'hypothèse de l'effet de dilution. Toutefois, les différences observées entre sites ne sont pas significatives malgré le grand nombre de spécimens analysés, ce qui démontre une nouvelle fois la difficulté de ce type d'étude. La collecte d'échantillons dans des localités où la circulation parasitaire est plus intense pourrait permettre l'observation de tendances plus franches. D'une manière générale, l'extension spatio-temporelle de l'étude est nécessaire avant d'aboutir à des conclusions solides.

Ce travail a permis d'apporter de, premiers éléments éco-épidémiologiques intéressants tout en validant l'approche utilisée avant d'entreprendre une étude de plus grande envergure. Plus de 20.000 phlébotomes supplémentaires ont déjà été collectés dans différentes localités en Guyane et sont en cours d'analyse. Par ailleurs, les outils moléculaires développés pourront être réemployés dans le cadre d'autres projets à venir et devraient ainsi contribuer à progresser dans la connaissance des leishmanioses au delà de cette thèse.

TITLE:

Biodiversity and infectious diseases: Impact of human activities on the sylvatic transmission cycle of leishmaniases in French Guiana.

ABSTRACT:

The dilution effect hypothesis states that more diverse ecological communities are less prone to pathogen transmission because of the presence of non-competent hosts acting as epidemiological dead-ends. In this work, we investigate the existence of this phenomenon in the case of zoonotic cutaneous leishmaniases in French Guiana. Molecular tools based on high-throughput sequencing technologies have been developed to study the epidemiological system. These tools were employed to explore leishmaniases transmission cycles in forest sites undergoing different levels of human-induced perturbations. Our results seem generally congruent with the dilution effect hypothesis, indicating higher disease risk in the most perturbed site. However, differences observed between sites were not significant, and more data is needed to draw general conclusions

AUTEUR : Arthur Kocher

TITRE :

Biodiversité et maladies infectieuses : Impact des activités humaines sur le cycle de transmission des leishmanioses en Guyane.

DIRECTEURS DE THÈSE :

Jérôme Murienne et Anne-Laure Bañuls

LIEU ET DATE DE SOUTENANCE :

Université Paul Sabatier, Toulouse, le 23 juin 2017

RÉSUMÉ :

Selon l'hypothèse de l'effet de dilution, les écosystèmes les plus riches en espèces seraient également les moins propices à la circulation des agents infectieux du fait de la présence d'hôtes non compétents constituant des impasses épidémiologiques. Dans cette thèse, nous explorons l'existence de ce phénomène sur le modèle des leishmanioses cutanées zoonotiques en Guyane. Des outils moléculaires basés sur l'utilisation des technologies de séquençage haut-débit ont été développés pour étudier le système épidémiologique. Ces outils ont ensuite été employés pour caractériser le cycle de transmission des leishmanioses dans des sites forestiers sujets à différents niveaux de perturbation d'origine humaine. Nos résultats semblent globalement congruents avec l'hypothèse de l'effet de dilution, et indiquent un risque infectieux plus élevé sur le site le plus perturbé. Toutefois, les différences observées entre les sites ne sont pas significatives et d'avantage de données seraient nécessaires pour tirer des conclusions générales.

MOTS-CLÉS : Écologie, effet de dilution, metabarcoding, Guyane

DISCIPLINE ADMINISTRATIVE : Écologie

INTITULÉ ET ADRESSE DU LABORATOIRE : Laboratoire Évolution & Diversité Biologique (EDB) - UMR 5174 (CNRS/UPS/ENFA), Université Paul SABATIER, Bâtiment 4R1. 118 route de Narbonne, 31062 Toulouse cedex 9, FRANCE